

MASTER'S THESIS

**Leiden aanpassingen in de PGA-notatie tot een verbeterde intuïtiviteit?
Een experiment in een praktische bedrijfscontext**

Tinnemans, Y.

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

Open Universiteit
www.ou.nl



Leiden aanpassingen in de PGA-notatie tot een verbeterde intuïtiviteit?

Een experiment in een praktische bedrijfscontext.

Do adjustments to the PGA-notation lead to an improved intuitivity?

An experiment in a practical business context.

Opleiding:	Open Universiteit, faculteit Bètawetenschappen
	Masteropleiding Business Process Management & IT
Cursus:	IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Yvonne Tinnemans
Identiteitsnummer:	
Datum:	25-06-2021
Afstudeerbegeleider	dr. Ben Roelens
Meelezer	
Derde beoordelaar	
Versie nummer:	1
Status:	definitief

Abstract

Enterprise Architecture modellen maken organisaties door middel van een visueel model inzichtelijk. Er zijn verschillende modelleertalen, waaronder *Process-Goal Alignment* (PGA). De PGA-modeleermethode is een specifieke *Domain Specific Modeling Language* (DSML). De PGA-taal heeft als doel om in de bedrijfsarchitectuur een goede strategische afstemming te bereiken. In een eerder onderzoek is deze modelleertaal onderzocht op intuïtiviteit. Met intuïtiviteit wordt de mate bedoeld waarin het element zonder verdere uitleg kan worden begrepen en gebruikt. In dit eerdere onderzoek is de aanbeveling gedaan om de taal meer intuïtief te maken door diverse elementen in de taal aan te passen. In dit onderzoek is deze aangepaste notatie in een daarvoor ontwikkeld experiment vergeleken met de oorspronkelijke notatie om te onderzoeken of de aanpassing heeft geleid tot een hogere intuïtiviteit van de taal. Op basis van de resultaten van dit onderzoek kan worden gesteld dat de aangepaste elementen van de PGA-notatie een hogere intuïtiviteit hebben. Er kan niet worden geconcludeerd dat de nieuwe elementen een hogere doeltreffendheid hebben. In de resultaten per element kan op basis van dit onderzoek worden geconcludeerd dat bepaalde elementen in de aangepaste notatie meer intuïtief zijn dan de oorspronkelijke notatie maar dat er ook elementen zijn die gelijk of minder intuïtief zijn in de nieuwe notatie ten opzichte van de oorspronkelijke notatie.

Sleutelbegrippen

Modelleren, modelleertaal, PGA, Process Goal Alignment, intuïtiviteit, doeltreffendheid.

Samenvatting

Enterprise Architecture (EA) geeft op hoog abstractieniveau een overzicht van bedrijfsprocessen en IT-systemen, de onderlinge relaties en gemeenschappelijk gebruik hiervan door de verschillende bedrijfsonderdelen. De *Unified Modeling Language* (UML) is op dit moment de standaardtaal om softwaresystemen te specificeren, visualiseren, construeren en documenteren (Lankhorst, 2009). UML is één van de *General Purpose Modeling Languages* (GPMLs) die worden gebruikt. Een GPML biedt een algemene weergave van zaken waarbij deze weergave in diverse contexten kan worden gebruikt. Er bestaan ook domeinspecifieke modeleertalen; *Domain-Specific Modeling Languages* (DSMLs). De *Process-Goal Alignment* (PGA) modelleermethode is een specifieke DSML. PGA betreft een modelleertaal die als doel heeft om in de bedrijfsarchitectuur een goede strategische afstemming te bereiken. Het is belangrijk dat deze taal intuïtief is. 'Intuïtiviteit van de notatie heeft betrekking op de mate waarin de visuele weergave zonder verdere uitleg kan worden begrepen en worden gebruikt' (Roelens & Bork, 2020). Daarmee zou een gebruiker de taal direct moeten kunnen begrijpen en gebruiken. De intuïtiviteit van de notatie is van belang omdat deze het eerste contact van de gebruiker is met de taal en daarmee ook een eerste voorwaarde is voor het aannemen en correct gebruiken daarvan. Door een intuïtieve taal te ontwikkelen, hoeft de gebruiker minder inspanningen te verrichten om de taal te begrijpen en te gebruiken.

In een eerder onderzoek is de intuïtiviteit van deze modelleertaal onderzocht en is er op basis van de resultaten een aangepaste notatie ontwikkeld door de onderzoekers waardoor de intuïtiviteit van de taal verbeterd zou kunnen zijn. De aanpassingen in de taal waren nog niet eerder geëvalueerd. Met dit onderzoek zijn, door gebruik te maken van een experiment, de aanpassingen in de taal in een praktische bedrijfscontext geëvalueerd. Hierbij is de volgende probleemstelling geformuleerd: '*In welke mate en op welke manier leiden de aanpassingen in de PGA-notatie en daarmee de nieuwe notatie tot een significant betere intuïtiviteit dan de oorspronkelijke notatie?*'

De afhankelijke variabelen in dit onderzoek zijn intuïtiviteit en doeltreffendheid. De intuïtiviteit is meetbaar gemaakt door het juiste antwoord bij de notaties te koppelen. Daarna is de (mate van) juistheid gekoppeld aan de tijd die de respondent nodig heeft gehad om tot dit antwoord te komen. De doeltreffendheid is de verhouding van de effectiviteit (het aantal juiste antwoorden dat door de respondenten wordt gegeven op de gestelde vragen) en efficiëntie (de hoeveelheid tijd die er nodig is om tot dit antwoord te komen). Voor het experiment zijn volgende alternatieve hypotheses geformuleerd:

Hypothese H_{a_I} : *De intuïtiviteit van PGA-notatie II is hoger dan die van PGA-notatie I*

Hypothese H_{a_D} : *De doeltreffendheid van PGA-notatie II is hoger dan die van PGA-notatie I*

Om de probleemstelling en de hypotheses te onderzoeken is er een experimentele techniek ontwikkeld om de aanpassing in de notatie te evalueren. Het experiment bestaat uit diverse vraagvormen die zijn voorgelegd aan de respondenten. Er is gebruik gemaakt van notatieassociatie in twee verschillende vormen. In het eerste onderdeel kregen de respondenten telkens een variant te zien van één van de zes gewijzigde elementen (klantdoel, intern doel, competentie, belangrijkheid, waardepropositie en waardeestroom); in totaal 12 elementen. Dit waren open vragen waarbij de respondenten de mogelijkheid hadden om voor ieder element drie termen te noemen die zij associeerden met de elementen. In het tweede onderdeel werden dezelfde 12 elementen gebruikt waarbij zij bij elke notatie de keuze hadden uit 4 antwoordmogelijkheden. Daarna zijn er begripsvragen bij een bedrijfsarchitectuur in PGA-taal voorgelegd aan de respondenten. Als laatste onderdeel van het experiment is de respondenten voorgelegd welke element op basis van de definitie het meest intuïtief vinden (perceptievragen). Er hebben in totaal 15 respondenten het experiment afgerond. De resultaten zijn geanalyseerd en op basis van deze analyse kan worden

gesteld dat de aangepaste PGA-notatie meer intuïtief is dan de oorspronkelijke PGA-notatie. Er is niet aangetoond dat de aangepaste PGA-notatie meer doeltreffend is dan de oorspronkelijke PGA-notatie. In de individuele resultaten kan op basis van dit onderzoek worden geconcludeerd dat bepaalde elementen in de aangepaste notatie meer intuïtief zijn dan de oorspronkelijke notatie, maar ook dat er elementen zijn die minder of gelijk intuïtief zijn ten opzichte van de oorspronkelijke notatie.

Summary

Enterprise Architecture (EA) provides a high abstraction level overview of business processes and IT systems, their mutual relationships and common use by the various business units. The Unified Modeling Language (UML) is currently the standard language for specifying, visualizing, constructing, and documenting software systems (Lankhorst, 2009). UML is one of the General-Purpose Modeling Languages (GPMLs) that are used. A GPML provides a general view of things where this view can be used in various contexts. There are also domain-specific modeling languages; Domain-Specific Modeling Languages (DSMLs). The Process-Goal Alignment (PGA) modeling method is a specific DSML. PGA is a modeling language that aims to achieve good strategic alignment in business architecture. It is important that this language is intuitive. 'Intuitiveness of the notation relates to the extent to which the visual representation can be understood and used without further explanation' (Roelens & Bork, 2020). This should allow a user to understand and use the language immediately. The intuitiveness of the notation is important because it is the user's first contact with the language and is therefore also a prerequisite for its adoption and correct use. Developing an intuitive language reduces the effort of the user to understand and use the language.

In a previous study, the intuitiveness of this modeling language was investigated and based on the results, a modified notation was developed by the researchers, which could improve the intuitiveness of the language. The changes in the language had not yet been evaluated. With this research, using an experiment, the adaptations in the language in a practical business context were evaluated. The following problem statement was formulated: 'To what extent and in what way do the adjustments in the PGA notation and thus the new notation lead to significantly better intuitiveness than the original notation?'

The dependent variables in this study are intuitiveness and effectiveness. The intuitiveness has been made measurable by linking the correct answer to the notations. The (degree of) correctness is then linked to the time it took the respondent to arrive at this answer. Effectiveness is the ratio of effectiveness (the number of correct answers given by the respondents to the questions asked) and efficiency (the amount of time it takes to arrive at this answer). The following alternative hypotheses were formulated for the experiment:

Hypothesis Ha_I: *The intuitiveness of PGA notation II is higher than that of PGA notation I*

Hypothesis Ha_D: *The effectiveness of PGA notation II is higher than that of PGA notation I*

To investigate the problem statement and the hypotheses, an experimental technique has been developed to evaluate the adaptation in the notation. The experiment consists of various question forms that were presented to the respondents. Notation association has been used in two different forms. In the first part, the respondents were shown a variant of one of the six changed elements (customer goal, internal goal, competence, importance, value proposition and value stream); a total of 12 elements. These were open-ended questions in which respondents had the option to name three terms for each view that they associated with the notation. In the second part, the same 12 notations were used. Respondents had a choice of 4 answer options for each notation. After that, understanding questions for a business architecture in PGA language were presented to the respondents. As a final part of the experiment, the respondents were asked which element they found to be the most intuitive based on the definition (perception questions). A total of 15 respondents completed the experiment. The results have been analyzed and based on this analysis it can be stated that the modified PGA notation is more intuitive than the original PGA notation. The modified PGA notation has not been shown to be more effective than the original PGA notation. In the individual results, it can be concluded on the basis of this research that certain elements in the modified notation are more intuitive than the original notation, but also that there are elements that are less or equally intuitive compared to the original notation.

Inhoudsopgave

Samenvatting.....	iii
Summary	v
1. Introductie.....	1
1.1. Achtergrond.....	1
1.2. Gebiedsverkenning.....	1
1.3. Probleemstelling.....	2
1.4. Opdrachtformulering	2
1.5. Motivatie/ relevantie	3
1.6. Aanpak in hoofdlijnen	3
2. Theoretisch kader.....	4
2.1. Onderzoeksaanpak.....	4
2.2. Uitvoering.....	4
2.3. Resultaten en conclusies	6
2.4. Doel van het vervolgonderzoek	9
3. Methodologie	10
3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n).....	10
3.2. Technisch ontwerp: uitwerking van de methode	11
3.3. Gegevensanalyse.....	12
3.4. Validiteit, betrouwbaarheid en ethische aspecten	13
4. Resultaten	15
4.1. Uitvoer van het onderzoek.....	15
4.2. Resultaten van het onderzoek	15
4.2.1. Demografische gegevens.....	15
4.2.2. Intuïtiviteit.....	16
4.2.3. Doeltreffendheid	19
4.3. Resultaten samengevat.....	21
5. Discussie, conclusies en aanbevelingen	22
5.1. Discussie – reflectie	22
5.1.1. Intuïtiviteit	22
5.1.2. Doeltreffendheid	24
5.1.3. Reflectie.....	24
5.2. Conclusies.....	25
5.3. Aanbevelingen voor de praktijk	26
5.4. Aanbeveling voor verder onderzoek	26
Bronnen.....	27

Bijlage 1: Literatuuranalyse	28
Bijlage 2: Vragenlijst en gebruikte modellen in de enquête.	31
Bijlage 3: Bedrijfsarchitectuur begripsvragen	32
Bijlage 4: Data-bestand	36

1. Introductie

1.1. Achtergrond

Enterprise Architecture (EA) kan worden omschreven als een overzicht op hoog abstractieniveau van de bedrijfsprocessen en IT-systemen van een onderneming, de onderlinge relaties en de mate waarin deze processen en systemen worden gedeeld door verschillende delen van de onderneming (Tamm et al., 2011). EA kan ook omschreven worden vanuit een huidige (*AS-IS*) en toekomstig gewenste toestand (*TO-BE*) van de processen en mogelijkheden en IT-infrastructuur van een organisatie. Daarnaast biedt het een *roadmap* om dit toekomstig doel te bereiken vanuit de huidige toestand (Shanks et al., 2018). Om de structuur van een organisatie vast te leggen wordt er onder andere gebruik gemaakt van documenten en visuele modellen (zgn. artefacten). Een verzameling van deze artefacten is volgens EA-professionals voldoende om een beeld te vormen van de organisatie en een geïntegreerd EA-model te vormen. EA-modellen illustreren de omschrijving van de organisatie en zij maken de organisatie door middel van een visueel model inzichtelijk. Professionals en onderzoekers hebben een groot aantal modelleertechnieken ontworpen (zie paragraaf 1.2) die variëren op het gebied van abstractieniveau, lagen in de organisatie, grafische notatie, enzovoort.

1.2. Gebiedsverkenning

De *Unified Modeling Language* (UML) is op dit moment de standaardtaal om softwaresystemen te specificeren, visualiseren, construeren en documenteren (Lankhorst, 2009). UML is één van de *General Purpose Modeling Languages* (GPMLs) die worden gebruikt. Een GPML biedt een algemene weergave van zaken waarbij deze weergave in diverse contexten kan worden gebruikt. Naast deze GPMLs bestaan ook domein-specifieke modelleertalen; *Domain-Specific Modeling Languages* (DSMLs). Deze DSMLs winnen populariteit omdat deze speciaal zijn ontworpen voor een bepaald doel en/of domein en omdat deze de complexiteit van het modeleren verminderen door concepten te gebruiken die bekend zijn bij de eindgebruikers (Roelens & Bork, 2020).

De *Process-Goal Alignment* (PGA) modelleermethode is een specifieke DSML. Deze heeft als doel om in de bedrijfsarchitectuur een goede strategische afstemming te bereiken (Roelens & Bork, 2020). Deze afstemming is een belangrijk doel in de organisatie omdat dit ervoor zorgt dat de bedrijfsstrategie overeenkomt met de interne infrastructuur en de processen. Dit zorgt ervoor dat bedrijven adequaat kunnen reageren op kansen en bedreigingen in de externe omgeving. Eén van de ontwerpvereisten van de PGA-methode is een duidelijke communicatie van de organisatiestrategie om ervoor te zorgen dat deze wordt begrepen door bedrijfsgerichte experts. Deze experts zouden bijvoorbeeld procesmanagers kunnen zijn, die kennis hebben van wat de eindgebruikers nodig hebben, maar die ook samenspraak hebben met een modelleur die het proces weergeeft. Deze experts gebruiken de methode niet zelf, maar worden geleid door de modelleur. De modelleur verzamelt de benodigde informatie en ontwerpt de PGA-modellen. Het is daarom van het grootste belang om ervoor te zorgen dat PGA-modellen intuïtief kunnen worden begrepen door de bedrijfsgerichte eindgebruikers om de cognitieve belasting voor hen te verminderen (Roelens & Bork, 2020). Bork, Schruffer en Karagiannis (2019, p. 2) omschrijven intuïtiviteit als volgt: 'Intuïtiviteit van de notatie heeft betrekking op de mate waarin de visuele weergave zonder verdere uitleg kan worden begrepen en worden gebruikt.' In de eerste blik kan de lezer daarmee de notatie makkelijk begrijpen en interpreteren. De intuïtiviteit van de notatie is van belang omdat deze het eerste contact van de gebruiker is met de taal en daarmee ook een eerste voorwaarde is voor het aannemen en correct gebruiken ervan. Een notatie moet daarom de modelleur helpen bij het maken en de gebruiker bij het interpreteren van een model. Daarbij zouden een aantal cognitieve taken verschuiven naar perceptuele taken waarbij bijvoorbeeld kennis minder belangrijk zal worden en het interpreteren en opvatten van hetgeen is omschreven belangrijker zal worden (Bork, Schruffer, & Karagiannis, 2019).

In eerder onderzoek van Roelens en Bork (2020) is de intuïtiviteit van de PGA-notatie geëvalueerd in drie fases. Na een korte introductie, is door de onderzoekers in de eerste fase (i.e. *Term Association*) gevraagd aan de respondenten om op basis van een genoemde term één of meerdere intuïtieve grafische weergaves van de term op te stellen en te tekenen. In de tweede fase (i.e. *Notation Association*) zijn de bestaande grafische weergaves van de PGA-notatie aan de respondenten voorgelegd. Aan de hand hiervan hebben zij ieder drie intuïtieve termen genoteerd die bij hen opkwamen zodra ze deze notatie zagen. In de derde fase (i.e. *Case Study Experiment*) is een *heatmap* aan de respondenten voorgelegd zonder de betekenis van de diverse weergaven aan te geven. Aan de hand hiervan moesten de respondenten vragen beantwoorden. Om deze vragen te kunnen beantwoorden moesten zij de *heatmap* kunnen interpreteren. Tenslotte is de respondenten nog gevraagd om feedback te geven op de notatie en eventuele verbeteringsuggesties te doen. Analyse van deze evaluatie heeft geleid tot een aangepaste notatie (Roelens & Bork, 2020).

1.3. Probleemstelling

Uit het eerdere onderzoek van Roelens en Bork (2020) blijkt dat de intuïtiviteit van de PGA-notatie kon worden verbeterd. Er zijn op basis van een evaluatiemethode verbetervoorstellen gedaan om de bestaande notatie aan te passen. De generaliseerbaarheid van het onderzoek lijkt enigszins beperkt. Het onderzoek heeft immers plaatsgevonden onder een groep van Masterstudenten en niet onder respondenten die werkzaam zijn in het betreffende bedrijfsdomein. Daarbij wordt verondersteld dat medewerkers uit het domein betere aanbevelingen hadden kunnen doen. Desondanks, bleek de evaluatietechniek bruikbaar en leverde verbeteringen op in de notatie (Roelens & Bork, 2020). De voorgestelde verbeteringen zijn echter afhankelijk van de beoordeling van de onderzoekers. De verbeteringen in de notatie zijn voorgesteld, maar er is niet verder onderzocht of de intuïtiviteit van de nieuwe notatie significant beter is dan de eerdere notatie. Dit leidt tot de volgende hoofdvraag:

‘In welke mate en op welke manier leiden de aanpassingen in de PGA-notatie en daarmee de nieuwe notatie tot een significant betere intuïtiviteit dan de oorspronkelijke notatie?’

1.4. Opdrachtformulering

Om bovengenoemde probleemstelling te gaan onderzoeken, is het belangrijk om te weten hoe de intuïtiviteit van de taal gemeten kan worden. Middels een literatuuronderzoek is onderzocht welke methode geschikt is om het verschil in intuïtiviteit tussen de oude en de nieuwe notatie te gaan onderzoeken. Aan de hand van de reeds bestaande methoden die in de literatuur is achterhaald, is een nieuwe methode opgezet om de intuïtiviteit te beoordelen door middel van een experiment.

Als eerste is de volgende deelvraag onderzocht: *‘Op welke manier kan de intuïtiviteit van een taal worden beoordeeld en gemeten?’*. Op deze manier zal inzichtelijk worden gemaakt wat van belang is om de betreffende modeleertaal op intuïtiviteit te beoordelen, wat er nodig is om dit te beoordelen en hoe kan dit meetbaar worden gemaakt. Dit blijkt uit een literatuuronderzoek. De tweede deelvraag heeft betrekking op het ontwikkelen van een methode om de intuïtiviteit van de twee talen met elkaar te gaan vergelijken. Daarbij wordt de volgende deelvraag beantwoord: *‘Op welke wijze kunnen de twee omschreven PGA-talen het best worden vergeleken voor wat betreft intuïtiviteit.’*. Op basis van het beantwoorden van de eerste en tweede deelvraag kan een derde deelvraag worden geformuleerd om het onderzoek in de praktijk uit te voeren middels een experiment. De derde deelvraag luidt: *‘Hoe wordt de intuïtiviteit van de beide notaties vergeleken tijdens het uitvoeren van een experiment?’*.

Gezien de huidige COVID-19-situatie zal het experiment op afstand moeten kunnen worden uitgevoerd.

1.5. Motivatie/ relevantie

Het onderzoek is relevant en draagt bij aan de *body of knowledge* vanwege de toenemende populariteit van DSMLs en de behoefte om deze beter af te stemmen op de behoefte van de makers en gebruikers. De populariteit van DSMLs neemt toe omdat deze vanwege de afstemming op het domein beter te begrijpen zijn door de eindgebruikers (Dalgarno & Fowler, 2008). Een andere reden van de toenemende populariteit en daarmee ook een reden om bovenstaande verder te onderzoeken is het voordeel van de DSML bij het gebruik van eventuele tools. Volgens Dalgarno en Fowler (2008) zijn in een tool die voor GPMLs kunnen worden gebruikt niet-relevante onderdelen vaak moeilijk te verwijderen. Eventueel te gebruiken diagrammen zijn niet afgestemd op het specifieke gebruik en hebben door afwijkende semantiek niet de juiste lading. Een voordeel van de DSML-benadering is dat de modelleeromgeving de 'taal' van het domein kan beperken en valideren. Dit is moeilijk met een GPML-profiel. Als voordeel voor het gebruik van een GPML wordt benoemd dat hier lagere kosten mee gemoeid zijn. Dit is te wijten aan het feit dat er voor GPML al een basis beschikbaar is, welke eventueel nog moet worden aangepast. Voor de DSML moet dit vanaf de grond worden opgebouwd (Dalgarno & Fowler, 2008).

Het gebruik van DSMLs zorgt voor een afnemende complexiteit en zorgt ervoor dat relevante aspecten goed worden weergegeven. Dit leidt tot een betere begrijpbaarheid van de modellen en tot een betere communicatie tussen mensen. Om dit te bereiken is er volgens Bork, Schruffer en Karagiannis (2019) meer kennis nodig met betrekking tot intuïtieve notaties en de manieren om deze te ontwikkelen en te evalueren. Op dit moment wordt er in de literatuur voornamelijk gesproken over de begrijpbaarheid van een bepaalde taal. Meer interessant en relevant is wellicht de intuïtiviteit van de notatie, waarover in paragraaf 1.2 al een en ander is toegelicht, omdat deze meer betrekking heeft op gevoelsmatigheid/het natuurlijk begrip van een bepaalde notatie. Dit aspect is nog maar weinig onderzocht in de bestaande wetenschappelijke literatuur (Bork, Schruffer, & Karagiannis, 2019). Door de intuïtiviteit van de notatie te verbeteren kan de gebruiksvriendelijkheid verbeterd worden, zonder gebruik te maken van extra uitleg bij de visuele weergave. Door een experiment uit te voeren met betrekking tot de intuïtiviteit van een bepaalde 'taal' zal hierover meer bekend worden. Daarnaast is door dit onderzoek een experimentele methode worden opgezet om een DSML op intuïtiviteit te evalueren. Een juiste evaluatie van de taal en daarmee het aanpassen van gebruik van een bepaalde notatie leidt naar verwachting tot een snellere en betere interpretatie en verwerking van de informatie. Het werken met deze notaties zou dan op een meer natuurlijke manier kunnen gaan waardoor de EA-communicatie makkelijker zou kunnen verlopen. Volgens Bork, Schruffer en Karagiannis (2019) is dit een *gap* die verder onderzocht dient te worden. De intuïtiviteit meten van een GPML is problematisch vanwege de verschillende aspecten van de 'taal'. Als je echter een DSML zou willen evalueren op intuïtiviteit zou dit beter mogelijk moeten zijn omdat potentiële gebruikers en de doelen van het gebruik van de 'taal' onderdeel van het ontwerp gaan uitmaken (Bork, Schruffer, & Karagiannis, 2019). De onderdelen van de taal en de meetbaarheid hiervan zal in hoofdstuk 2 verder worden toegelicht.

1.6. Aanpak in hoofdlijnen

In hoofdstuk 2 van dit rapport zal worden stilgestaan bij het literatuuronderzoek dat heeft plaatsgevonden om tot het opzetten van een model voor de toetsing van de intuïtiviteit en de vergelijking van de beide notaties te komen. In hoofdstuk 3 zal worden ingegaan op de opzet van het experiment om de beide notaties met elkaar te vergelijken en te waarden op gebied van intuïtiviteit. De resultaten van het onderzoek worden in hoofdstuk 4 weergegeven. In hoofdstuk 5 wordt verder ingegaan op de resultaten en is de conclusie en aanbeveling voor de praktijk en aanbeveling tot verder onderzoek te vinden.

2. Theoretisch kader

In hoofdstuk 1 is een introductie gegeven met betrekking tot de modelleertalen en meer specifiek de PGA-modelleertaal en het onderzoek van Roelens en Bork (2020) waarin een voorstel is gedaan om de intuïtiviteit van deze taal te verbeteren. In dit theoretisch kader wordt onderzocht welke methodes er vanuit de literatuur relevant zijn om de intuïtiviteit van notaties vergelijkend empirisch te evalueren.

2.1. Onderzoeksaanpak

Door in het literatuuronderzoek gebruik te maken van een *integrative review* (Saunders, Lewis, & Thornhill, 2016) is vanuit diverse invalshoeken in kaart gebracht hoe het antwoord op de eerste deelvraag kan worden gevonden. Er is daarmee antwoord gezocht op: *‘Op welke manier kan de intuïtiviteit van een taal worden beoordeeld en gemeten?’*. Daarmee is inzichtelijk gemaakt wat van belang is om de betreffende modelleertaal op intuïtiviteit te kunnen beoordelen. Om deze deelvraag te kunnen beantwoorden is van belang om te weten wat intuïtiviteit precies inhoudt, maar is ook van belang om te onderzoeken hoe de intuïtiviteit kan worden beoordeeld en wat daarvoor nodig is en hoe de intuïtiviteit van twee modelleertalen kan worden vergeleken.

Na raadplegen van het artikel van Roelens en Bork (2020) is er behoefte om vanuit de literatuur verdieping te vinden in de bovengenoemde zaken. Vanuit deze publicatie is dan ook het literatuuronderzoek gestart. Het literatuuronderzoek is uitgevoerd in de periode tussen september 2020 en december 2020 in de digitale universiteitsbibliotheek van de Open Universiteit. Daarbij is gebruik gemaakt van de *snowballing*-techniek. De gebruikte query's en selectiecriteria zijn te vinden in tabel 1. In deze tabel zijn bij de query's, waarvoor dit relevant is, ook de stopcriteria opgenomen. Soms is er direct naar een artikel gezocht en daarmee is het zoeken gestopt zodra het betreffende artikel is gevonden. In andere situatie is met behulp van diverse criteria bepaald om het zoeken met deze betreffende gegevens te stoppen. Dit kan zijn op basis van een top-5 selectie, met Web Of Science citaties of omdat de volgende artikelen uit een ander werkveld lijken te komen. Voor de diverse query's zijn het aantal onderzochte papers dan ook telkens verschillend. Over de selectie van de betreffende artikelen wordt in de volgende paragraaf meer toegelicht.

2.2. Uitvoering

Uit onderstaande tabel (tabel 1) blijken de resultaten per gebruikte query en ook welke selectie daaruit is gemaakt. In dit hoofdstuk zal worden omschreven hoe de selectie in zijn werk is gegaan, welke andere artikelen zijn geanalyseerd en waarom deze selectie van de artikelen is gemaakt.

Tabel 1: Query's

Nr.	Query	Publicatiedatum	Gesorteerd op	Aantal resultaten	Stopcriterium	Aantal relevante artikelen	Taal
1.	Burton-Jones	2009	Relevantie	3691	Gestopt na beoogde artikel gevonden	1 (Burton-Jones, Wand, & Weber, 2009)	Engels
2.	'empirical evaluation' en 'conceptual modeling'	Geen	Relevantie	94.583	Gestopt na ontbreken <i>empirical evaluation</i> in titel	2 (Gemino & Wand, 2004; Wand & Weber, 2002)	Engels
3.	'empirical evaluation' en 'conceptual modeling'	Vanaf 2005	Relevantie	82.071	Gestopt na 5 artikelen zonder relevantie. Web of Science citaties lager dan 50.	0	Engels

4.	Snowballing eerder gevonden artikelen	Vanuit relevante artikelen uit query 1 en 2	Doornemen van de artikelen	Zie referenties betreffende artikel	Gestopt nadat Web of Science citaties lager was dan 100.	2 (Burton-Jones & Meso, 2006; Gemino & Wand, 2005)	Engels
5.	'Understanding' en 'Business Process Models'	Vanaf 2015	Relevantie	552.055	Gestopt nadat Web of Science citaties drastisch afnam.	1 (Figl, 2017)	Engels
6.	Snowballing eerder gevonden artikel	Vanuit relevant artikel query 5	Doornemen van de artikelen	Zie referenties betreffende artikel	Gestopt nadat Web of Science citaties daalden en de abstract van eerste 5 gevonden artikelen niet meer relevant was.	1 (Figl & Recker, 2016)	Engels
7.	Intuitiveness	Geen	Relevantie	6.641	Gestopt nadat onderzoeksgebied niet meer 'Information Technologie' was	1 (Turner, 2008)	Engels
8.	Intuitiveness en understanding	Geen	Relevantie	3.372	Gestopt nadat onderwerp onderzoeksgebied niet meer 'human factors in computing systems' bevat.	1 (Grandhi, Joue, & Mittelberg, 2011)	Engels

De publicatie van Roelens en Bork (2020) geeft daarbij het startpunt om aanvullende literatuur te raadplegen. Vanuit de referenties van deze publicatie is gezocht naar relevante publicaties voor het vervolgonderzoek. De publicatie van Burton-Jones, Wand en Weber (2009) is volledig geraadpleegd vanwege de aanwezigheid van de kernbegrippen en is ook relevant gebleken voor het verder onderzoek vanwege het gebruik in de uitgangspublicatie en verdere beschikbare aanwezige informatie inzake het onderwerp. Analyse van andere artikelen heeft plaatsgevonden verder in het zoekproces naar aanleiding van het relevante artikel. Vanuit het artikel van (Burton-Jones, Wand, & Weber, 2009) blijken vanuit de titel al de twee kernbegrippen voor verder onderzoek; namelijk *empirical evaluation* en *conceptual modeling*. Deze hebben namelijk betrekking op de informatiebehoefte leidend tot het onderzoek.

De tweede query is hier dan ook op gebaseerd. Hierbij is gebruik gemaakt van dezelfde twee kernbegrippen en een sortering op relevantie om niet te veel resultaten uit te sluiten. Vanuit de top 5 van de resultaten hebben 2 artikelen het grootste aantal verwijzingen binnen 'Web of Science'. Op basis hiervan zijn de artikelen van Gemino en Wand (2004) en Wand en Weber (2002) geselecteerd. Er is gekozen voor de 'top 5' van de resultaten omdat buiten deze resultaten *empirical evaluation* niet meer in de titel, abstract en de inleiding van de resultaten voorkomen. De overige artikelen uit de 'top 5' zijn op basis van de lagere score dan de geselecteerde artikelen voor wat betreft referenties door 'Web of Science' doorgenomen, maar niet meer relevant gebleken en daarom ook niet verder betrokken in het onderzoek. Omdat de twee geselecteerde artikelen al wat ouder zijn; 2002 en 2004, is besloten om een publicatiedatum toe te voegen bij de selectiecriteria om zo meer recente informatie te verkrijgen. De datum van 2005 is daarbij gekozen omdat dit net ligt na de publicatie van de laatst gevonden artikelen. Uit deze query kwamen een kleiner aantal resultaten, maar vanwege de lage relevantie en het lage aantal 'Web of Science'-citaties is besloten om geen resultaten uit deze query te betrekken in de onderzoeksresultaten.

De gevonden publicaties zijn geanalyseerd en er bleek nog behoefte aan aanvullende informatie. Daarom is uit de referenties van het artikel van Burton-Jones, Wand en Weber (2009) verder gezocht en zijn Burton-Jones en Meso (2006) en Gemino en Wand (2005) geselecteerd. Uit de artikelen Wand en Weber (2002) en Gemino en Wand (2004) zijn geen referenties meer geselecteerd omdat de artikelen al ouder zijn en er een grotere behoefte ontstond aan meer recente informatie.

Op basis hiervan is dan ook gezocht naar artikelen met een meer recente publicatiedatum (< 5 jaar geleden) en daarnaast ontbreekt er nog informatie over het begrijpen van de modellen. Daarbij is query 5 uitgevoerd en één relevant artikel geselecteerd Figl (2017). Op basis van dit artikel is in de referenties van dit artikel gezocht naar meer relevante artikelen over het 'begrijpen' en de gewenste eigenschappen van de gebruiker van de modellen. Op basis hiervan is artikel Figl en Recker (2016) geselecteerd als relevante literatuur. Daarmee was aannemelijk dat er voorlopig genoeg informatie beschikbaar over modellen en de begrijpbaarheid hiervan.

Op dat moment ontbrak nog literatuur met betrekking tot intuïtiviteit. Zoals eerder in de paragraaf omschreven, was er nog behoefte aan informatie over wat dit inhoudt en hoe de intuïtiviteit gemeten kan worden. Om deze reden is er gezocht op intuïtiviteit en gekeken naar de relevantie van het artikel voor wat betreft het onderzoeksgebied waar dit betrekking op heeft. Het eerste resultaat was de meest relevante omdat deze paper betrekking heeft op *Information Technologie*, terwijl de volgende artikelen een meer medische achtergrond hadden. Na het lezen van de abstract van de eerste 5 artikelen is daarom enkel het eerste artikel van Turner (2008) meegenomen in de resultaten. Op basis hiervan zou er nog een koppeling gelegd moeten worden tussen hoe intuïtiviteit en begrijpbaarheid samenhangen in een dergelijke situatie. Zo is query 8 ontstaan waarbij ook enkel het eerste resultaat, het artikel van Grandhi, Joue en Mittelberg (2011), meegenomen is als relevante literatuur. Dit artikel werd als enige relevant geacht omdat dit als onderzoeksgebied de menselijke factoren op het gebruik en ontwerp van computersystemen heeft. De overige resultaten hadden een meer medische achtergrond; zo bleek na analyse van de abstract en onderzoeksgebied. Dit was het moment waarop de verzamelde onderwerpen compleet leken en diepgaand zijn geanalyseerd.

2.3. Resultaten en conclusies

Naar aanleiding van de analyse van de papers is in de bijlage (zie bijlage 1) een korte samenvatting per artikel uitgewerkt. In tabel 2 (zie volgende pagina) zijn de kenmerken van de artikelen samengevat. Daarbij zijn het jaar van publicatie, het type artikel, het doel van het onderzoek, de maatregelen, het experimentele ontwerp, het gegeven of zij informatie bevatten over intuïtiviteit, welke type vragen zijn gesteld en welke respondenten betrokken waren in het onderzoek. Voor het type vragen is er onderscheid gemaakt tussen het volgende type vragen die in eerder onderzoek van Gemino en Wand (2004) zijn benoemd:

- *Recall (R) vs Comprehension (C)*. Dit is een begrippentest waarbij vragen worden gesteld over modelementen. Dit kan plaatsvinden op basis van de inhoud van het model (i.e. comprehension) of nadat het model weggenomen is (i.e. recall).
- *Problem solving (P)*: Deze test focust op het daadwerkelijke begrijpen van het domein en niet op het begrijpen van de individuele onderdelen van het model. Door de deelnemer een 'probleem' te laten oplossen wordt er niet alleen naar het diagram zelf gekeken, maar naar het grotere geheel en de betekenis.
- *Cloze (CI)*. In dit geval wordt de deelnemers gevraagd om blanco plekken aan te vullen naar aanleiding van een omschrijving van het betrokken domein.

Tabel 2: Resultaten

Paper	Auteur	Publicatie-jaar	Web of Science	Gebruikt model	Methodologie/experiment	Afhankelijke variabele	Experimenteel design	Domein of modelbegrip	Operationele procedures	Voorwaarden	Intuïtiviteit	Type vragen	Respondenten
1.	(Gemino & Wand, 2005)	2004	J	ERM	Experiment	Niveau van domeinbegrip door de gebruiker	Between subjects	Model	<ul style="list-style-type: none"> • 12 mp (R) • 5 open vragen (P) • 45 invul (CI) 	Min. één Management Information System cursus gevolgd	N	R/ P/ CI	77 studenten
2.	(Burton-Jones & Meso, 2006)	2006	J	GDM UML	Experiment	Feitelijk en waargenomen begrip	Between subjects	Model	<ul style="list-style-type: none"> • 11 open vragen (P) 35 minuten • 5 invul (CI) 25 minuten 	Object-Oriented Design-student + Object-Oriented Analysis cursus	N	P/ CI	57 studenten
3.	(Figl, 2017)	2017	J	div.	Methodologie	objectieve begripsnauwkeurigheid tijdgebonden begripbaarheid moeilijkheid van subjectief begrip en gebruikersvoorkeur	Empirisch onderzoek review	Domein	<ul style="list-style-type: none"> • 76 geanalyseerde papers 	Empirische studies die de begripbaarheid van procesmodellen objectief meten + Empirische studies die de gebruikers-voorkeur en de begripbaarheid van procesmodellen subjectief meten + Theoretische discussies over de begripbaarheid van procesmodellen.	J	n.v.t.	literatuur
4.	(Figl & Recker, 2016)	2014	J	ERM UML BPMN ORM	Experiment	Tekst vs. gestructureerde tekst Tekst vs. diagram Gestructureerde tekst vs. diagram Iconen vs. geen iconen	Within subjects	Domein	<ul style="list-style-type: none"> • Information cover sheet Sectie met vragen: Demografie Ervaring met conceptuele modellen Cognitieve stijlen Verschillende procesweergaves Taakspecifieke voorkeuren 	Gecontroleerde omgeving + Business school studenten + vloeiend Engels en Duits moedertaal	J	R	120 studenten
5.	(Gemino & Wand, 2004)	2004	J	CMT	Methodologie	Opstellen van een framework om conceptuele modellen empirisch te evalueren.	Empirisch onderzoek review	Model	n.v.t.	n.v.t.	N	CI	literatuur
6.	(Burton-Jones, Wand, & Weber, 2009)	2009	N	CMT	Methodologie	Geven van richtlijnen om conceptuele modellen te evalueren.	Empirisch onderzoek review	Model	28 potentiële studies ter evaluatie	n.v.t.	N	CI	literatuur
7.	(Turner, 2008)	2007	J	n.v.t.	Methodologie	Definiëren van intuïtiviteit	Empirisch onderzoek review	Domein	n.v.t.	n.v.t.	J	CI	literatuur
8.	(Grandhi, Joue, & Mittelberg, 2011)	2011	N	n.v.t.	Experiment	Intuïtiviteit van gebaren	Within subjects	Domein	<ul style="list-style-type: none"> • 19 scenario's • 3 conditions (instructional, habitual, instructional using body-part-as-tool) 	Opname video Aan onderzoeker worden gebaren uitgebeeld door deelnemers	J	R	16 businessusers
9.	(Wand & Weber, 2002)	2002	J	n.v.t.	Methodologie	Ontwikkeling van richtlijnen om te modelleren	Empirisch onderzoek review	Domein	n.v.t.	n.v.t.	N	CI	literatuur

Uit de literatuur blijkt dat er op diverse gebieden kansen zijn tot mogelijk nader onderzoek. De meest relevante artikelen geven informatie over de begrijpbaarheid van de gebruikte modellen. De artikelen zoals die van Gemino en Wand (2005) en van Burton-Jones en Meso (2006) spreken over de begrijpbaarheid van een bepaald model en niet over de intuïtiviteit. Er is weinig literatuur beschikbaar over intuïtiviteit en dan meer specifiek op het gebied van de modelleertalen. Het is belangrijk de intuïtiviteit in kaart te brengen omdat dit het maken, lezen en gebruiken van de modellen verbetert en vergemakkelijkt. Naast de beperkte literatuur met betrekking tot onderzoek naar intuïtiviteit in de modellen valt in de literatuurstudie ook op dat de meeste experimenten zoals de artikelen van Gemino en Wand (2005), van Burton-Jones en Meso (2006) en van (Figl, 2017) worden uitgevoerd met studenten. Experimenten waarbij businessusers van de modelleertaal zijn betrokken zijn (bijna) niet gevonden. Enkel is het artikel van Grandhi, Joue en Mittelberg (2011) gevonden waarbij businessusers zijn gebruikt in het experiment. Vaak wordt in de onderzoeken gebruik gemaakt van groepen studenten vanwege de uitvoerbaarheid van het onderzoek. De modellen worden juist in de praktijk gebruikt en de mate van intuïtiviteit in de praktijk is daarmee van belang. Kennis op dit gebied ontbreekt op dit moment nog in de literatuur.

De eerste deelvraag vanuit de literatuurstudie luidt '*Op welke manier kan de intuïtiviteit van een taal worden beoordeeld en gemeten?*'. Om de intuïtiviteit van een modelleertaal te beoordelen is belangrijk te kijken wat het meest natuurlijk voelt voor de gebruiker. Eerder ging het over de begrijpbaarheid van een bepaalde taal, maar de intuïtiviteit heeft betrekking op het op een natuurlijke wijze kunnen gebruiken van de taal. In de geanalyseerde literatuur zijn diverse onderzoeken opgezet om aan de hand van de antwoorden van de deelnemers de begrijpbaarheid van de taal te beoordelen. Zo is in het onderzoek van Figl en Recker (2016) diverse type vragen aan de respondenten voorgelegd. Het eerste onderdeel betrof gegevens van de respondent zelf: leeftijd, geslacht, opleidingsniveau en of er ervaring is met procesmodelleren. In het tweede onderdeel van het onderzoek werden er een aantal modellen bij de respondenten voorgelegd waarbij de respondenten dienden aan te geven wat voor type model het was op basis van meerkeuzeopties (5 opties en 'ik weet het niet'). Als derde onderdeel is er een vragenlijst met een vijfpunt Likertschaal voorgelegd uit het paper van Blazhenkova en Kozhevnikov (2009). In deze vragenlijst werden 45 stellingen gepresenteerd met onder andere betrekking tot de voorkeuren van tekstuele uitleg of voorkeur tot een afbeelding. Als vierde zijn er zes visuele presentaties voorgelegd bij de respondenten welke zij ieder 50 seconden mochten bestuderen. Later volgde een tekst waarin deze moeten worden ingevuld. Tenslotte werden er vragen gesteld met betrekking tot de voorkeur voor een bepaald notatie van de respondenten (Figl & Recker, 2016). In het onderzoek van Grandhi, Joue en Mittelberg (2011) is een experiment opgezet door 16 deelnemers (*businessusers*) 19 acties op drie verschillende manieren te laten tonen aan de onderzoekers: 1. *instructional*; waarbij de deelnemer de onderzoeker instrueert wat te doen, 2. *habitual*; waarbij de deelnemer aan de onderzoeker uitlegt hoe hij/zij dit normaal zouden doen, 3. *instructional using, body-part-as-tool*; waarbij de deelnemer de *Instructional* herhaalt maar waarbij zij het 'gereedschap' middels hun hand uitbeeldden. Daarna is de deelnemers een reactie gevraagd met betrekking tot de uit te beelden gebaren en wat hen daarbij opviel. Daarbij wordt de intuïtiviteit en het gemak van de handgebaren verder bevestigd door opmerkingen van de deelnemers dat bepaalde wijze van gebaren niet als natuurlijk voelt omdat zij zoiets normaal gesproken niet zouden doen.

In de diverse experimenten die zijn uitgevoerd, wordt gebruik gemaakt van verschillende type vraagstellingen zoals Gemino en Wand (2004) deze eerder hebben gebruikt. Door selectie (bijvoorbeeld op beroepsgroep en voorkennis van modelleertalen) van respondenten en de hoeveelheid informatie die zij voorafgaand aan het experiment krijgen, kan er worden gestuurd om onderscheid te maken tussen begripbaarheid en intuïtiviteit. Als bepaalde respondenten meer ervaring hebben vanuit functie of voorkennis van het weergegeven proces bestaat de mogelijkheid dat zij hetgeen beter begrijpen. Dit geeft echter geen antwoord op de mate van intuïtiviteit van de elementen. De geanalyseerde paper van Grandhi, Joue en Mittelberg (2011) maakt gebruik van een manier om de notaties of gebaren te vergelijken waarbij de respondenten aangeven wat zij als meest goed ervaren en daarmee intuïtief zijn. De intuïtiviteit van een taal kan worden beoordeeld door op verschillende manieren aan de deelnemers de diverse vraagtypen voor te leggen om op deze manier te concluderen welke van de twee notaties meer intuïtief is. In de paper van Gemino en Wand (2005) is gebruik gemaakt van *recall*, *comprehension*, *problemsolving* en *cloze* vragen en ook evaluatievragen te stellen aan de respondenten gericht op hoe intuïtief zij bepaalde onderdelen van de notatie ervaren. Door gebruik te maken van verschillende vraagstellingen kan naar verwachting een eenduidige conclusie worden getrokken over de intuïtiviteit van de notaties. Daarbij is ook de snelheid waarmee de respondent kan antwoorden van belang. Kort samengevat kan daarmee worden gesteld dat de juiste interpretatie van de notatie en de tijd die de respondent hiervoor nodig heeft, de indicatoren zijn voor het meten van de intuïtiviteit en doeltreffendheid. Wanneer de juiste interpretatie van de notatie in een kortere tijd kan worden gevonden wordt hiermee verondersteld dat de intuïtiviteit en doeltreffendheid van deze notatie beter is dan die van de andere notatie. Hiermee wordt de eerste deelvraag beantwoord.

2.4. Doel van het vervolgonderzoek

Zoals in de vorige paragraaf omschreven ontbreekt er in de onderzochte literatuur, kennis op het gebied van intuïtiviteit als ook resultaten van onderzoeken die uitgevoerd zijn bij *businessusers*. Doel van het vervolgonderzoek is het beantwoorden van de tweede en derde deelvraag:

- ‘Op welke wijze kunnen de twee omschreven PGA-talen het best worden vergeleken voor wat betreft intuïtiviteit.’
- ‘Hoe wordt de intuïtiviteit van de beide notaties vergeleken tijdens het uitvoeren van een experiment?’

Deze deelvragen zullen door middel van het opzetten van een experimentele methode, om de intuïtiviteit van de twee PGA-notaties te evalueren en te meten of deze doeltreffend is, worden beantwoord. Er wordt daarbij gekozen voor een praktische bedrijfscontext om zo *businessusers* te gebruiken als respondenten voor het onderzoek. Dit specifiek omdat de eindgebruikers in de *businessusers* ook de daadwerkelijke gebruikers zijn van de PGA-techniek en omdat de meeste onderzoeken worden uitgevoerd onder studenten en er daarmee kennis ontbreekt met betrekking tot de *businessusers*. In het volgende hoofdstuk (3) zal worden ingegaan op het opzetten van een experiment waarbij er naar verwachting kennis zal worden opgedaan aangaande deze twee gevonden *gaps*.

3. Methodologie

In het vorige hoofdstuk is omschreven welke relevante literatuur is onderzocht en op welke manier de intuïtiviteit kan worden geëvalueerd. In dit hoofdstuk zal verder worden ingegaan op de wijze van opzetten van het vervolgonderzoek.

3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

De experimentele techniek die is opgezet door Bork, Schruffer en Karagiannis (2019) heeft als doel om een meer intuïtieve notatie tot stand te brengen. Er is echter nog niet vastgesteld of de aanpassing uit de paper van Roelens en Bork (2020) heeft geleid tot een verbetering voor wat betreft intuïtiviteit. Daarnaast is de nieuwe notatie voor een deel afhankelijk van de menselijke creativiteit van de onderzoekers en de respondenten en er is daarmee ook nog onvoldoende geëvalueerd of dit een betere notatie is voor wat betreft intuïtiviteit. Dit leidt tot de tweede deelvraag binnen dit onderzoek:

‘Op welke wijze kunnen de twee omschreven PGA-talen het best worden vergeleken voor wat betreft intuïtiviteit?’

Om de vraag in de probleemstelling te kunnen beantwoorden is er een methode ontwikkeld om de beide notaties met elkaar te kunnen vergelijken. Er is gekozen voor een experiment in een *within-subjects design* (Saunders, Lewis, & Thornhill, 2016). Er zijn verschillende onderzoeksmethoden met ieder zijn eigen voor- en nadelen. Er worden in dit experiment diverse variabelen tegen elkaar afgezet. Daarbij is onderscheid gemaakt tussen de onafhankelijke en afhankelijke variabelen (Saunders, Lewis, & Thornhill, 2016). Het uitvoeren van een experiment is het meest geschikt omdat het doel van het onderzoek is om (in een gecontroleerde omgeving) na te gaan of het veranderen van de onafhankelijke variabele (i.e. een verschillende versie van de PGA-notatie) leidt tot een verandering of verbetering in, in dit geval, de intuïtiviteit en de doeltreffendheid (de afhankelijke variabelen) (Saunders, Lewis, & Thornhill, 2016). Het experiment is op basis van de verkregen inzichten uit hoofdstuk 2 worden opgesteld.

Andere onderzoeksmethoden als een survey of case study zijn hier vanwege de kenmerken en uitvoerbaarheid minder geschikt voor. Een survey is bijvoorbeeld minder geschikt vanwege het type vraagstellingen dat er wordt gebruikt. (Saunders, Lewis, & Thornhill, 2016). Doel van het onderzoek is om te testen of de notatie intuïtief is. Dit resultaat kan middels een survey moeilijker worden gemeten. Een case study is minder geschikt vanwege de uitvoerbaarheid in een bepaalde setting (Saunders, Lewis, & Thornhill, 2016). De controle met betrekking tot hetgeen er wordt voorgelegd is dan niet of minder aanwezig en de meetbaarheid van de intuïtiviteit is daarmee minder. Er is niet gekozen voor een quasi-experiment omdat er geen sprake is van een experimentele en een controlegroep. Hiervan zou bijvoorbeeld gebruik gemaakt worden als respondenten in een bepaalde groep (wel of niet controlegroep of als team) ingedeeld moeten worden (Saunders, Lewis, & Thornhill, 2016). Dit is in deze situatie niet het geval en dus is het uitvoeren van een quasi-experiment niet wenselijk. Het splitsen van de groep respondenten in een reguliere groep en controlegroep is eveneens niet wenselijk omdat het experiment zal worden uitgevoerd onder een relatief klein aantal respondenten en daarmee de groep te klein zou worden om persoonlijke kenmerken van de respondenten in balans te brengen. De groep respondenten is wel verdeeld in een tweetal groepen op basis van de geboortedatum van de respondent. Dit heeft te maken met het toetsen van de twee modellen in de nieuwe en oude notatie en om op deze manier het leereffect te minimaliseren en toch een vergelijking te kunnen maken op basis van de begripsvragen.

Door de opgezette methode in te zetten in een praktische bedrijfscontext en voor te leggen aan de geselecteerde respondenten zijn de resultaten verzameld. Er is specifiek gekozen voor een praktische bedrijfscontext en daarmee voor businessusers omdat vanuit de literatuur blijkt dat er meerdere malen onderzoek is verricht bij studenten, maar niet genoeg bij businessusers. Daarmee wordt er gewerkt aan één van de *gaps* die blijkt uit de literatuur uit hoofdstuk 2. Na uitvoeren van het

experiment zijn de resultaten geanalyseerd om tot beantwoording van de hoofdvraag te komen. Hieruit blijkt dan of kan worden gesteld dat de aangepaste notatie inderdaad meer intuïtief en doeltreffend is dan de oorspronkelijke notatie. Het onderzoek heeft daarmee een deductieve aanpak, omdat er vanuit de literatuur een onderzoek is opgezet om de theorie te testen (Saunders, Lewis, & Thornhill, 2016). Vanuit de theorie zijn hypotheses geformuleerd welke naar aanleiding van de dataverzameling zijn getest.

3.2. Technisch ontwerp: uitwerking van de methode

Om het experiment vorm te kunnen geven dienden er variabelen geformuleerd te worden. De onafhankelijke variabelen zijn daarbij de twee versies van de PGA-notatie, waarbij er in het experiment wordt gemeten of de aangepaste notatie (i.e. notatie II) meer intuïtief is dan de oorspronkelijke notatie (i.e. notatie I) en daarmee meer doeltreffend is. De afhankelijke variabelen zijn intuïtiviteit en doeltreffendheid. Deze afhankelijke variabelen zijn gekozen op basis van de aanbevelingen gedaan in eerder onderzoek van Roelens en Bork (2020) en Bork, Schruffer en Karagiannis (2019). 'Intuïtiviteit van de notatie heeft betrekking op de mate waarin de visuele weergave zonder verdere uitleg kan worden begrepen en worden gebruikt.' (Bork, Schruffer, & Karagiannis, 2019, p. 2). Dit zal meetbaar worden gemaakt door het aantal juiste antwoorden bij de notaties te meten. Daarnaast zullen de respondenten perceptievragen worden voorgelegd die eveneens worden meegenomen in de beoordeling voor de mate van de intuïtiviteit. De doeltreffendheid is de verhouding van de effectiviteit (het aantal juiste antwoorden dat door de respondenten wordt gegeven op de gestelde vragen) op de efficiëntie (de hoeveelheid tijd die er nodig is om tot dit antwoord te komen). Er wordt dus onderzocht of notatie II meer intuïtief is dan notatie I en of er door het hanteren van de nieuwe notatie een verbetering is voor wat betreft doeltreffendheid. In het experiment zullen per afhankelijke variabele twee hypotheses worden geformuleerd; een nulhypothese en een alternatieve hypothese (Saunders, Lewis, & Thornhill, 2016).

Hypothese H_{0_I} : *Er is geen verschil in intuïtiviteit tussen PGA-notatie I en II*

Hypothese H_{a_I} : *De intuïtiviteit van PGA-notatie II is hoger dan die van PGA-notatie I*

Hypothese H_{0_D} : *Er is geen verschil in doeltreffendheid tussen PGA-notatie I en II*

Hypothese H_{a_D} : *De doeltreffendheid van PGA-notatie II is hoger dan die van PGA-notatie I*

Vanwege het aantal beschikbare respondenten (i.e. ongeveer 10) is ervoor gekozen om een *within-subjects design* te kiezen. Er zijn zo immers minder respondenten nodig om het experiment goed te kunnen uitvoeren want de groepen zouden bij een *between-subjects design* te klein worden. De resultaten zijn dan wellicht minder statistisch betrouwbaar. Bij een *within-subjects design* is er slechts één groep respondenten. Daarmee worden aan alle respondenten beide notaties voorgelegd, zonder dat er sprake is van een controlegroep.

Alvorens het onderzoek verder kon gaan diende er een vragenlijst ontwikkeld te worden waarmee de geformuleerde hypotheses onderzocht kunnen worden. De vragenlijst is ontwikkeld aan de hand van de eerder geraadpleegde literatuur.

Voor de start van het inhoudelijke experiment wordt de respondent de toestemmingsverklaring getoond (zie bijlage 3) waarbij zij een keuze moeten maken of zij hiermee akkoord gaan. Mocht de respondent hiermee niet akkoord gaan wordt het experiment beëindigd. Er is eerst een korte introductie gegeven over modelleertalen waarbij geen vermelding is gemaakt van het doel van het experiment of de gebruikte modelleertaal. Om ervoor te zorgen dat het kennisniveau en domein van de respondenten op een vergelijkbaar niveau ligt, is ervoor gekozen om het experiment uit te voeren onder bankmedewerkers die werkzaam zijn op een IT-gerelateerde afdeling. Daarnaast is onderzoeker in dit domein actief en is er om deze reden beter toegang tot deze respondenten. Na een korte inleiding van het experiment kregen de respondenten een vragenlijst voorgelegd die uit diverse onderdelen bestaat. Het eerste deel bestaat uit een aantal vragen die betrekking hebben op

persoonskenmerken. Zo kan achteraf achterhaald worden of deze persoonskenmerken geen verstorende invloed hebben binnen het kader van de geformuleerde hypothesen. Daarnaast is het geboortjaar van de respondent gebruikt om te bepalen welke versie van het experiment zal worden voorgelegd. De vragen zullen er echter niet toe leiden dat respondenten als persoon te identificeren zijn. In bijlage 2 is de gehele vragenlijst beschikbaar. Er is gevraagd naar geslacht, geboortedatum, hoogst genoten opleiding, sector waarin respondent werkzaam is, op welk niveau de respondent werkzaam is, of er ervaring is met modelleren en de inschatting van de eigen modelleerkennis. Daarna volgen de inhoudelijke vragen welke bestaat uit diverse soorten vragen. Uit de literatuur in hoofdstuk 2 blijken de verschillende soorten vragen. Om een zo'n compleet mogelijk beeld te vormen, hebben de respondenten *notation association* vragen gekregen die onder andere door Burton-Jones en Meso (2006) en Gemino en Wand (2005) in hun onderzoek zijn gebruikt. De *notation association* is in twee varianten voorgelegd. In eerste gedeelte notatieassociatie kregen de respondenten de 12 elementen afzonderlijk te zien waarbij zij door middel van open tekstvelden de associatie met het element (tot 3 antwoorden mogelijk) konden noteren. Daarna werd in notatie associatie (2) dezelfde elementen apart bevraagd, maar dit keer als meerkeuzevragen met vier antwoordopties. Daarnaast zijn er *comprehension questions*; 'begripsvragen' gesteld zoals die in hun onderzoek ook door Gemino en Wand (2005) worden benoemd. Hierbij zijn twee omschreven bedrijfsarchitecturen (bijlage 3) die zowel in de oorspronkelijk als in de aangepaste notatie zijn voorgelegd. Hierbij werden op basis van dit figuur vragen gesteld zodat de respondenten een bepaald 'probleem' dienen op te lossen. Om de invloed van het ontwikkelen van inzicht tijdens deelname aan het experiment (een leereffect) op te heffen, is ervoor gekozen om de twee versies van de bedrijfsarchitecturen in een verschillende volgorde voor te leggen aan de respondenten. De groep met een oneven geboortjaar heeft van model 1 de oude notatie en van model 2 de nieuwe notatie voorgelegd gekregen. De groep met een even geboortjaar heeft van model 1 de nieuwe notatie voorgelegd gekregen en van model 2 de oude notatie (bijlage 3). De begripsvragen zijn ook in een bepaalde volgorde gesteld om zo min mogelijk informatie prijs te geven inzake de betekenis van de notaties. Als laatste onderdeel is respondenten de vraag voorgelegd hoe intuïtief zij bepaalde notaties ervaren; de perceptievragen. Hierbij wordt eerst aangegeven wat er onder intuïtief wordt verstaan en dan worden de twee dezelfde notaties in de oorspronkelijke en nieuwe variant getoond waarbij de respondent kan aangeven welke hij het meest intuïtief vindt. Dit kan door middel van een schuifbalk, waarbij 0 de linkse notatie en 100 de rechtse notatie betreft. Een score van 50 geeft aan dat zij beiden even of juist niet intuïtief worden ervaren. Er is ruimte voor opmerkingen om de keuze te onderbouwen. In het onderzoek van Grandhi, Joue en Mittelberg (2011) is dit type vraag ook gesteld aan de respondenten. Na dit onderdeel werd het experiment afgesloten.

Er is gedurende het experiment bewaakt dat eerdere antwoorden niet meer kunnen worden aangepast zodat opgedane kennis tijdens het experiment niet gebruikt kan worden om antwoorden uit voorgaande taken aan te passen. Het leereffect wordt daarmee verminderd. Het opstellen van de vragenlijst dient daarmee rekening te houden. Het experiment zal in verband met de huidige coronasituatie op afstand uitgevoerd moeten worden door middel van Limesurvey. Via deze tool kan naast registratie van de gegeven antwoorden ook de reactietijd van de respondenten worden gemeten.

3.3. Gegevensanalyse

Na het uitvoeren van het experiment zijn er data beschikbaar. Deze zal op een juiste manier verwerkt moeten worden om tot een betrouwbare conclusie van het onderzoek te kunnen komen. De verzamelde data bestaan uit de antwoorden van de respondenten en ook de antwoordtijd van de respondenten (categorische en intervaldata). Voor de notatieassociatie geldt dat er een kader zal moeten worden geschetst waarbinnen de antwoorden van de respondenten moeten vallen om juist te zijn. Op het moment dat het antwoord volledig fout is zijn er 0 punten toegekend, bij het deels raken van de betekenis is er één punt toegekend en bij het volledig juist hebben van het antwoord is er 2 punten toegekend. Bij notatieassociatie is er telkens sprake van één correct antwoord welke vooraf is

vastgelegd. Ditzelfde geldt voor de begripsvragen. De data met betrekking tot de antwoordtijd (intervalvariabele) zijn na het uitvoeren van het experiment beschikbaar binnen Limesurvey en zijn gebruikt om de efficiëntie van het experiment te bepalen. Om de doeltreffendheid te bepalen zijn effectiviteit en efficiëntie tegen elkaar afgezet. Hierbij is de mate van effectiviteit net zo belangrijk als de mate van efficiëntie waardoor zij allebei een weging '1' krijgen om de doeltreffendheid te bepalen.

Op statistische wijze is gemeten of de resultaten en de gevonden relaties zich al dan niet berusten op waarschijnlijkheid. Hiermee wordt een grenswaarde gesteld op basis waarvan de nulhypothese kan worden verworpen. Hiertoe kan door middel van 'statistische analyse de significantie (of p-waarde) worden bepaald. Hiervoor is de vuistregel dat deze waarde kleiner dient te zijn dan 0,05 om de nulhypothese te kunnen verwerpen (Saunders, Lewis, & Thornhill, 2016). Deze significantie (of de p-waarde) is een getal tussen de 0 en 1, waarmee in toenemende mate wordt aangegeven in welke mate de gevonden relaties zich op willekeur bevinden. Hoe dichter de waarde bij 0 ligt, hoe meer aangenomen mag worden dat de gevonden resultaten een werkelijke associatie aanduidt. De genoemde drempelwaarde van 0,05 betekent dat de kans dat de gevonden resultaten zich op toeval baseert, 5% is (Saunders, Lewis, & Thornhill, 2016).

Allereerst zijn de resultaten die voor de statistische analyse relevant zijn geïdentificeerd. Het gaat hierbij naast de demografische variabelen, ook om de scores en benodigde tijd voor het beantwoorden van de vragen over notatieassociatie en notatieassociatie 2, de perceptievragen, en de begripsvragen, waarmee de intuïtiviteit kan worden geanalyseerd. Voor de analyse van de doeltreffendheid is een berekening uitgevoerd waaruit blijkt hoeveel tijd men nodig heeft gehad om één scorepunt te behalen. Een lagere uitkomst veronderstelt dan een betere doeltreffendheid.

De variabelen zijn in SPSS allereerst geanalyseerd om normaalverdeling te bepalen of te weerleggen, doormiddel van de Shapiro-Wilk test. Dit is een vereiste omdat volgende stappen in de statistische analyse gebaseerd is op de aanname van normaal-verdeeldheid. Bij normaal verdeelde resultaten zijn gepaarde T-Tests uitgevoerd voor de scores en tijd van de notatieassociaties, en voor de gewogen score (tijd per scorepunt). Voor deze variabelen zijn bij niet-normaal verdeelde resultaten een exact sign test uitgevoerd. Voor de perceptievragen is bij normaal verdeelde resultaten een one-sample T-Test uitgevoerd, en bij niet-normaal verdeelde resultaten een One Sample Wilcoxon Signed Test. Een uitgebreide uitwerking hiervan is opgenomen in hoofdstuk 4.

3.4. Validiteit, betrouwbaarheid en ethische aspecten

Het is belangrijk dat de resultaten van het onderzoek betrouwbaar en valide zijn. De intuïtiviteit van de methodes wordt met elkaar vergeleken in het experiment. Dat dit ook daadwerkelijk wordt gemeten (constructvaliditeit), wordt beïnvloed door het type vragen die worden gesteld (Saunders, Lewis, & Thornhill, 2016). Het diverse type vragen die vanuit hoofdstuk 2 relevant blijken en zoals deze ook in hoofdstuk 2 en hoofdstuk 3.2 worden benoemd zorgen er tevens voor dat er een check plaatsvindt op de eerder gegeven antwoorden en dat daarmee het gegeven antwoord niet op toeval berust. Door diverse onderwerpen op verschillende manieren te bevragen worden de antwoorden van de respondenten gecheckt, gekoppeld aan eerdere antwoorden en worden er controlevragen gesteld aan de respondenten die ervoor zorgen dat de intuïtiviteit en de doeltreffendheid ook daadwerkelijk gemeten wordt. Door de notatie voor te leggen en de respondent zijn associatie erbij te laten noemen (notatieassociatie) en later in het experiment de respondenten te laten kiezen welke notatie er het best bij een bepaalde term past (perceptievragen) kan een controle worden uitgevoerd. Het is dan van belang dat de respondenten niet meer terug kunnen naar de eerder gegeven antwoorden om zo te voorkomen dat zij dit nog gaan aanpassen.

De interne validiteit wordt over het algemeen verhoogd door respondenten in een gecontroleerde omgeving te laten deelnemen aan het experiment (Saunders, Lewis, & Thornhill, 2016). De eerdere experimenten die zijn uitgevoerd werden veelal afgenomen bij studenten omdat dit zorgt voor een

controleerbare groep. Er is echter gekozen om het experiment uit te voeren binnen een praktische bedrijfscontext om de interne validiteit te verhogen omdat deze de respondenten de mogelijke gebruikers zijn van de notatie. Om de interne validiteit verder te verbeteren worden er respondenten betrokken met een profiel dat geschikt is in de context van de PGA-modelleertaal (IT'ers (grotendeels) binnen het bankwezen). Daarnaast worden zij aan het begin van het experiment voorzien van dezelfde informatie en wordt in het experiment nagegaan of zij ervaring hebben met modelleren. Door deze drie factoren mee te wegen in de resultaten worden de interne validiteit verbeterd.

De externe validiteit wordt gewaarborgd doordat de respondenten allemaal werkzaam zijn in eenzelfde sector. Het experiment kan daarmee na uitvoering gegeneraliseerd worden voor de IT-medewerkers.

Er wordt door een combinatie van verschillende vragen voor gezorgd dat de resultaten betrouwbaar zijn door het antwoord van meerdere kanten te laten bevestigen (i.e. triangulatie) (Saunders, Lewis, & Thornhill, 2016). Dit kan worden bereikt door de verschillende type vragen te stellen aan de respondenten. De verzamelde data en conclusies zullen daarnaast beoordeeld worden. Dit komt de betrouwbaarheid van de verzamelde data en de daaraan gekoppelde conclusies ten goede. Ook bestaat de mogelijkheid dat door triangulatie de resultaten minder eenduidig zijn, maar dit verhoogt de validiteit en betrouwbaarheid van het onderzoek (Thurmond, 2001).

De respondenten zijn en blijven anoniem gedurende en na het experiment en analyse van de resultaten. Er zijn geen vragen aan de respondenten gesteld die tot gevolg kunnen hebben dat zij aan de hand van de antwoorden te herleiden zijn tot een bepaald persoon. Er zijn ook geen vragen gesteld die specifieke bedrijfsinformatie zullen blootleggen. Deelname aan het experiment is op vrijwillige basis en er is expliciet om toestemming (toestemmingsverklaring) gevraagd om de gegevens vast te leggen. Op het moment dat respondenten niet meer wensen deel te nemen aan het experiment, hebben zij de mogelijkheid om op elk moment de deelname te beëindigen.

In hoofdstuk 4 zal verder worden ingegaan op het daadwerkelijk uitvoeren van het experiment en daarbij zal dan ook de derde deelvraag worden beantwoord: *'Hoe kan de intuïtiviteit van de beide notaties worden vergeleken door het uitvoeren van een experiment?'*.

4. Resultaten

In dit onderdeel zal worden besproken hoe het experiment is uitgevoerd en of daar eventueel afwijkingen zijn geconstateerd ten opzichte van hetgeen vooraf was bepaald. Daarna zullen de resultaten van het onderzoek worden weergegeven.

4.1. Uitvoer van het onderzoek

Het onderzoek is vanaf 27 april tot en met 17 mei 2021 beschikbaar geweest voor respondenten middels een enquête in limesurvey. In bijlage 2 is de enquête zichtbaar. In de enquête ontbreekt bij de begripsvragen het getoonde bedrijfsarchitecturen, maar deze is in bijlage 3 bijgevoegd. Gedurende de enquête was dit model te zien voor de respondenten. De potentiële respondenten waren vooraf aan het onderzoek door middel van een online oproep tot deelname in kaart gebracht zodat ook met een percentage dat niet zou deelnemen er genoeg respondenten zouden zijn voor deze vorm van het onderzoek. Gedurende het uitvoeren van het experiment is echter gebleken dat het aantrekken van voldoende respondenten met een IT-gerelateerde functie binnen de financiële dienstverlening niet geheel realiseerbaar was, omdat er uiteindelijk onvoldoende bereidheid was om deel te nemen aan het experiment. De link naar de enquête werd daarom openbaar gedeeld (met de vermelding van de doelgroep erbij) waardoor er een minder grote drempel was om deel te nemen aan het experiment. Dit werd gerealiseerd door activatie van een tweede enquête (met dezelfde vragen) in limesurvey zodat de resultaten apart verzameld konden worden. Het openbaar delen van de enquête heeft ertoe geleid dat er ook respondenten hebben deelgenomen die niet volledig aan het vooraf opgestelde profiel voldoen. Echter zijn de respondenten werkzaam in een IT-gerelateerde functie. De passende (IT-)functies en het tegenvallende aantal respondenten in de oorspronkelijke doelgroep (financiële dienstverlening) hebben er dus toe geleid dat het profiel van de respondenten is uitgebreid binnen het experiment.

4.2. Resultaten van het onderzoek

Er zijn in totaal 37 mensen gestart met het experiment waarvan 15 de enquête volledig hebben afgerond. De andere respondenten zijn allemaal gestopt voordat het inhoudelijke gedeelte van de enquête was gestart, waardoor hun responses niet zijn meegenomen in de resultaten.

4.2.1. Demografische gegevens

De vragen met betrekking de demografische gegevens zijn zichtbaar in de bijlage en geven een beknopt beeld van de respondent. In totaal namen 15 respondenten deel aan het onderzoek, waarvan 73.3% mannen ($N = 11$). De gemiddelde leeftijd van de respondenten is 38 jaar ($M = 1983$). Van de respondenten is 26.7% ($N = 4$) wetenschappelijk opgeleid, en 46.7% ($N = 7$) heeft een Hbo-opleiding afgerond, heeft 20% ($N = 3$) wel een Hbo- of WO-opleiding genoten maar deze niet afgerond, en heeft 1 respondent ($N = 1$) middelbaar onderwijs of vergelijkbaar afgerond (6.7%). 66.7% van de respondenten ($N = 10$) geeft aan ervaring te hebben met modeleren. Gemiddeld besteedde de respondenten 29.3 minuten ($M = 1758.56$, $SD = 594.01$) aan het experiment. De functie (en in eerste instantie ook de sector waarin de respondent werkzaam is) is belangrijk om te bepalen of de respondent geschikt is voor deelname aan het onderzoek. Daarnaast is het geboortjaar gebruikt om te bepalen welke versie van de begripsvragen de respondent voorgelegd kreeg (even vs. oneven).

4.2.2. Intuïtiviteit

Hieronder worden de resultaten besproken, waarmee het aspect intuïtiviteit van de twee versies van modelleertaal kunnen worden vergeleken. Eerst worden de statistische resultaten geïnterpreteerd per vraaggroep, t.w. notatieassociatie, notatieassociatie (2) en begripsvragen. Hieronder worden eerst de nul- en alternatieve hypothese herhaald, waarna per taak de analyse wordt uiteengezet.

Hypothese H_{0_I} : *Er is geen verschil in intuïtiviteit tussen PGA-notatie I en II*

Hypothese H_{a_I} : *De intuïtiviteit van PGA-notatie II is hoger dan die van PGA-notatie I*

Notatieassociatie- en begripsvragen

Door middel van de Shapiro-Wilk test is gekeken naar de normaalverdeling van de antwoorden gegeven op de notatieassociatievragen, vanwege de goede toepasbaarheid bij kleinere sample-groepen (Razali & Wah, 2011). Hierbij is de nulhypothese gevormd dat de variabelen normaal-verdeeld zijn, welke verworpen kan worden indien uit de Shapiro-Wilk test een significantieniveau van $p < 0.05$ voor de variabele blijkt: in dat geval zijn de variabelen niet-normaal verdeeld. Uit de analyse van de scores van de notatieassociatievragen blijkt dat de variabelen NA (2) nieuw ($W(15) = .92, p = .216$), en NA (2) oud ($W(15) = .90, p = .094$) normaal-verdeeld zijn; de nulhypothese is bevestigd. De variabelen NA nieuw ($W(15) = .72, p < .0005$), en NA nieuw ($W(15) = .28, p < .0005$) zijn niet-normaal verdeeld; de nulhypothese is verworpen voor deze variabelen.

Uit de analyse van de begripsvragen blijkt dat de variabele Totaal CQ OUD ($W(15) = .95, p = .567$) normaal-verdeeld is; de nulhypothese is bevestigd. De variabele Totaal CQ NIEUW ($W(15) = .86, p = .023$) is niet-normaal verdeeld; de nulhypothese is verworpen voor deze variabelen.

Notatieassociatie

Vanwege de niet-normaalverdeling voor de totaalscore van 'NA nieuw' en 'NA oud' is exact sign test uitgevoerd. De score voor de notatieassociatie laat een statistisch significante toename zien, bij toepassing van de nieuwe notatie ($p = .002$, rechtszijdig getoetst).

Frequencies

		N
Totaalscore NA nieuw - Totaalscore NA oud	Negative Differences ^a	0
	Positive Differences ^b	9
	Ties ^c	6
	Total	15

a. Totaalscore NA nieuw < Totaalscore NA oud

b. Totaalscore NA nieuw > Totaalscore NA oud

c. Totaalscore NA nieuw = Totaalscore NA oud

Notatieassociatie 2

Frequentieanalyse op de items waarmee de notatieassociatie (2) is gemeten gaf normaal-verdeelde resultaten. Voor elk van de presentatievoorkeuren is een gepaarde T-Test uitgevoerd. De totaalscore van de notatieassociatie 2 is met toepassing van het nieuwe model, vergeleken met het oude model, statistisch significant hoger, $t(14) = 3.833, p = .001$ (rechtszijdig getoetst).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Totaalscore NA (2) nieuw - Totaalscore NA (2) oud	1.267	1.280	.330	.558	1.975	3.833	14	.002

Begripsvragen

Vanwege de niet-normaalverdeling voor de scores voor de begripsvragen is exact sign test uitgevoerd. De score voor de begripsvragen laat een statistisch niet-significante afname zien, bij toepassing van de nieuwe notatie ($p = .605$, linkszijdig getoetst).

Frequencies

		N
Totaal CQ NIEUW - Totaal CQ OUD	Negative Differences ^a	8
	Positive Differences ^b	6
	Ties ^c	1
	Total	15

a. Totaal CQ NIEUW < Totaal CQ OUD

b. Totaal CQ NIEUW > Totaal CQ OUD

c. Totaal CQ NIEUW = Totaal CQ OUD

Perceptievragen

Door middel van de Shapiro-Wilk test is gekeken naar de normaalverdeling van de antwoorden gegeven op de perceptievragen, vanwege de goede toepasbaarheid bij kleinere sample-groepen (Razali & Wah, 2011). Hierbij is de nulhypothese gevormd dat de variabelen normaal-verdeeld zijn, welke verworpen kan worden indien uit de Shapiro-Wilk test een significantieniveau van $p < 0.05$ voor de variabele blijkt: in dat geval zijn de variabelen niet-normaal verdeeld. Uit de analyse van de perceptievragen blijkt dat de variabelen intern doel ($W(15) = .91$, $p = .136$), waardepropositie ($W(15) = .85$, $p = .017$), en klantdoel ($W(15) = .95$, $p = .555$) normaal-verdeeld zijn; de nulhypothese is bevestigd. De variabelen waardestroom ($W(15) = .82$, $p = .007$), en competentie ($W(15) = .86$, $p = .023$) zijn niet-normaal verdeeld; de nulhypothese is verworpen voor deze variabelen.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SPSS intern doel	.172	15	.200*	.910	15	.136
SPSS Waardepropositie	.190	15	.150	.850	15	.017
SPSS Waardestroom	.260	15	.007	.821	15	.007
SPSS Klantdoel	.143	15	.200*	.952	15	.555
SPSS Competentie	.213	15	.065	.859	15	.023

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Waardepropositie

Vanwege de normaal-verdeelde score voor het symbool waardepropositie, is een One-Sample T-Test uitgevoerd, met als testwaarde 50, ter aanduiding van geen voorkeur. De score van de presentatievoorkeur voor het symbool waardepropositie was statistisch niet-significant lager dan het middelpunt vo or de voorkeursaanduiding tussen de oude, en nieuwe notatie ($t(14) = -.614, p = .726$ (linkszijdig)).

One-Sample Test

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SPSS Waardepropositie	-.614	14	.549	-6.333	-28.45	15.78

Intern doel

Vanwege de normaal-verdeelde score voor het symbool intern doel, is een One-Sample T-Test uitgevoerd, met als testwaarde 50, ter aanduiding van geen voorkeur. De score van de presentatievoorkeur voor het symbool intern doel was statistisch niet-significant hoger dan het middelpunt voor de voorkeursaanduiding tussen de oude, en nieuwe notatie ($t(14) = .348, p = .367$ (rechtszijdig)).

One-Sample Test

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SPSS intern doel	.348	14	.733	3.000	-15.51	21.51

Competentie

De score van de presentatievoorkeur voor het symbool competentie was statistisch niet-significant lager dan het middelpunt voor de voorkeursaanduiding tussen de oude, en nieuwe notatie ($t(14) = -.454, p = .672$ (linkszijdig)).

One-Sample Test

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
SPSS Competentie	-.454	14	.657	-4.000	-22.90	14.90

Klantdoel

Vanwege de normaal-verdeelde score voor het symbool klantdoel, is een One-Sample T-Test uitgevoerd, met als testwaarde 50, ter aanduiding van geen voorkeur. De score van de presentatievoorkeur voor het symbool klantdoel was statistisch niet-significant hoger dan het middelpunt voor de voorkeursaanduiding tussen de oude, en nieuwe notatie ($t(14) = .414, p = .343$ (rechtszijdig)).

One-Sample Test

	t	df	Sig. (2-tailed)	Mean Difference	Test Value = 50 95% Confidence Interval of the Difference	
					Lower	Upper
SPSS Klantdoel	.414	14	.685	3.000	-12.55	18.55

Waardestroom

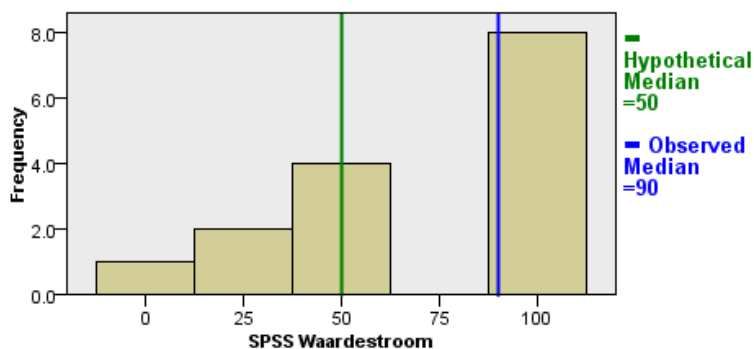
Vanwege de niet-normaalverdeling voor de score voor het symbool 'waardestroom' is een One Sample Wilcoxon Signed Test uitgevoerd. Hierbij werd gehypotheiseerd dat de mediaan van de score voor waardestroom 50 is, hetgeen aangeeft dat statistisch gezien geen voorkeur kan worden herleid voor de oude of nieuwe notatie. De analyse wees uit dat de nulhypothese niet verworpen kan worden ($p = .055$).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of SPSS Waardestroom equals 50.	One-Sample Wilcoxon Signed Rank Test	.055	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

One-Sample Wilcoxon Signed Rank Test



4.2.3. Doeltreffendheid

Hieronder worden de resultaten besproken, waarmee het aspect doeltreffendheid van de twee versies van modelleertaal kunnen worden vergeleken. Hieronder worden eerst de nul- en alternatieve hypothese herhaald, waarnaar per taak de analyse wordt uitgezet.

Hypothese H_{0_D} : *Er is geen verschil in doeltreffendheid tussen PGA-notatie I en II*

Hypothese H_{a_D} : *De doeltreffendheid van PGA-notatie II is hoger dan die van PGA-notatie I*

Tijd per scorepunt

Om het verschil in doeltreffendheid tussen de oude en nieuwe notatie inzichtelijk te maken, wordt een gewogen score berekend aan de hand van de behaalde score, en de tijd die respondenten nodig hadden om deze score te behalen.

De score is als volgt berekend: $\frac{\text{totale tijd}}{(1 + \text{behaalde score})}$. De behaalde score wordt met 1 opgehoogd, om delingen door nul te voorkomen. De uitkomst geeft hierdoor aan, hoeveel seconden nodig zijn geweest om een scorepunt te behalen, te rekenen vanaf 1 (tijd per scorepunt). Frequentieanalyse op verschillen in gewogen scores tussen de oude en nieuwe notatie is uitgevoerd.

Notatieassociatie

De tijd per scorepunt behaald bij het beantwoorden van de notatieassociatievragen (1) met toepassing van het nieuwe model, vergeleken met het oude model, is statistisch niet-significant lager, $t(14) = -2.294$, $p = .019$ (enkelzijdig).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	GewogenTijdScore NAnieuw - GewogenTijdScore NAoud	-43.85883	74.04545	19.11845	-84.86384	-2.85383	-2.294	14	.038

Notatieassociatie 2

De tijd per scorepunt behaald bij het beantwoorden van de notatieassociatievragen (2) met toepassing van het nieuwe model, vergeleken met het oude model, is statistisch significant lager, $t(14) = -3.344$, $p < .002$ (enkelzijdig): gemiddeld ruim 6 seconden per punt beter gescoord ($N = -6.24$).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	GewogenTijdScore NA2nieuw - GewogenTijdScore NA2oud	-6.23596	7.22319	1.86502	-10.23603	-2.23589	-3.344	14	.005

Begripsvragen

De gewogen score behaald bij het beantwoorden van de begripsvragen met toepassing van het nieuwe model, vergeleken met het oude model, is statistisch niet-significant hoger, $t(14) = -.192$, $p = .575$ (enkelzijdig).

Paired Samples Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	GewogenTijdScore CQnieuw - GewogenTijdScore CQoud	8.00515	161.30958	41.64995	-81.32511	97.33542	.192	14	.850

4.3. Resultaten samengevat

In onderstaande tabel worden de resultaten van de analyses kort herhaald, welke verder worden geïnterpreteerd in het volgende hoofdstuk.

Tabel 3: Resultaten

			<i>Resultaat</i>	<i>H₀ / H_a</i>
<i>Intuïtiviteit</i>	<i>Notatieassociatie- en begripsvragen</i>	Notatieassociatie	Toename, $p = .002$	H ₀ Verworpen H _a Aanvaard
		Notatieassociatie 2	$t(14) = 3.833, p = .001$	H ₀ Verworpen H _a Aanvaard
		Begripsvragen	Afname, $p = .605$	H ₀ Aanvaard H _a Verworpen
	<i>Perceptievragen</i>	Waardepropositie	$t(14) = -.614, p = .726$	H ₀ Aanvaard H _a Verworpen
		Intern doel	$t(14) = .348, p = .367$	H ₀ Aanvaard H _a Verworpen
		Competentie	$t(14) = -.454, p = .672$	H ₀ Aanvaard H _a Verworpen
		Klantdoel	$t(14) = .414, p = .343$	H ₀ Aanvaard H _a Verworpen
		Waardestroom	Aanvaard H ₀ , $p = .055$	H ₀ Aanvaard H _a Verworpen
<i>Doel- treffendheid</i>	<i>Notatieassociatie- en begripsvragen</i>	Notatieassociatie	$t(14) = -2.294, p = .019$	H ₀ Aanvaard H _a Verworpen
		Notatieassociatie 2	$t(14) = -3.344, p < .002$	H ₀ Verworpen H _a Verworpen
		Begripsvragen	$t(14) = -.192, p = .575$	H ₀ Aanvaard H _a Verworpen

5. Discussie, conclusies en aanbevelingen

De in hoofdstuk 4 gepresenteerde resultaten zullen in dit hoofdstuk verder worden besproken met daarbij de interpretatie. Verderop in het hoofdstuk volgt de conclusie en als laatste de aanbevelingen voor verder onderzoek.

5.1. Discussie – reflectie

In het onderzoek van Roelens en Bork (2020) is de bestaande PGA-notatie geëvalueerd. De respondenten van het onderzoek van Roelens en Bork (2020) betroffen Masterstudenten. Op basis van dat onderzoek is de notatie aangepast. De zes aangepaste elementen zijn in dit onderzoek geëvalueerd om te onderzoeken of het aanpassen van de notatie heeft geleid tot een betere intuïtiviteit van de taal en of er kan worden gesteld dat de aanpassing ook heeft geleid tot een betere doeltreffendheid. De probleemstelling die daarbij is geformuleerd is *‘In welke mate en op welke manier leiden de aanpassingen in de PGA-notatie en daarmee de nieuwe notatie tot een significant betere intuïtiviteit dan de oorspronkelijke notatie?’* Hieronder worden de resultaten van het onderzoek besproken.

5.1.1. Intuïtiviteit

Op basis van de gegeven antwoorden wordt bekeken of er voor de oude of de nieuwe notatie een betere score wordt gerealiseerd. Daarna zullen de antwoorden indien mogelijk per individueel element worden gebruikt om de score te verklaren.

Notatieassociatie

Voor de notatieassociatie, waarbij de respondenten dienden in te vullen welke woord of woorden er past bij de weergegeven notatie, is te zien dat de score voor wat betreft juiste antwoorden erg laag is. Het resultaat is wel statistisch significant en geeft aan dat de nieuwe notatie beter scoort dan de oude notatie. Dit valt te verklaren vanwege het feit dat voor één element van de nieuwe notatie (belangrijkheid) goed werd gescoord. Op basis hiervan is er statistisch gezien te concluderen dat de nieuwe notatie een betere intuïtiviteit heeft dan de oude notatie. Nuance die worden aangebracht betreft het feit dat één element een belangrijke invloed heeft op deze score.

Notatieassociatie (2)

Voor notatieassociatie (2), waarbij de respondenten op basis van vier antwoordmogelijkheden het voor hen meest passende begrip bij de notatie hebben gekozen, is de score (absoluut gezien) meer verdeeld. Statistische analyse, welke significant is, wijst uit dat de nieuwe notatie beter scoort dan de oude notatie. De analyses zijn uitgevoerd op notatie-niveau. Voor meer specifieke en gedetailleerde score per element, kan er ook worden gekeken naar de antwoorden per element. Absoluut gezien scoort het element voor Competentie, Intern doel, Waardestroom en Belangrijkheid inderdaad beter in de nieuwe notatie dan in de oude notatie. Echter scoort Waardepropositie slechter in de nieuwe notatie dan in de oude notatie. Voor Klantdoel is er geen verschil te zien tussen beide notaties.

Begripsvragen

Voor de begripsvragen, waarbij de respondenten op basis van een PGA-model gevraagd werd om de vragen met betrekking tot het model te beantwoorden, zijn geen statistisch significante resultaten gevonden.

Perceptievragen

Bij de perceptievragen is de betekenis van de elementen uitgelegd en hebben de respondenten zowel de oude als nieuwe notatie in beeld. Daarbij is hen gevraagd welke van de twee zij het meest intuïtief vinden, passend bij de gegeven uitleg. Respondenten konden middels een slider een waarde tussen 0 en 100. Hierbij betekende 0 de linkse notatie en 100 de rechtse notatie meer intuïtief; een score van 50 gaf aan dat beiden gelijk (niet) intuïtief zijn. Ook konden de respondenten aangeven waarom zij voor deze bepaalde score hebben gekozen en of zij eventueel beiden (niet) intuïtief vinden.

Waardepropositie

Voor de Waardepropositie werd aangegeven dat de oude notatie meer intuïtief is. Het resultaat is statistisch niet significant en er kan daarom geen eenduidige conclusie uit deze resultaten worden getrokken. Respondenten gaven bijna allemaal een aanvullende opmerking. Positief aan de nieuwe notatie werd genoemd 'het geven van iets (cadeau) legt het doel meer uit dan het evenwicht dat de weegschaal weergeeft' een andere respondent vindt het 'afwegen' dat door de weegschaal wordt uitgebeeld juist een betere weergave van hetgeen wordt bedoeld.

Intern doel

Voor het Intern doel werd aangegeven dat de nieuwe notatie meer intuïtief is. Het resultaat is statistisch niet significant en er kan daarom geen eenduidige conclusie uit deze resultaten worden getrokken. Eén van de aanvullende opmerkingen (welke in een vergelijkbare strekking meerdere keren werd genoemd) betreft: 'doordat het tandwiel in de fabriek wordt weergegeven, geeft het duidelijker aan dat het om iets gaat dat intern is. Anderen noemen de oude notatie beter omdat de nieuwe notatie meer lijkt aan te geven over een groter doel te gaan, over het gehele proces of over de gehele fabriek.

Competentie

Voor Competentie werd aangegeven dat de oude notatie meer intuïtief is. Het resultaat is statistisch niet significant en er kan daarom geen eenduidige conclusie uit deze resultaten worden getrokken. Vier van de vijftien respondenten geeft hierbij aan dat van de beiden elementen niet één specifiek beter is, met onder andere als reden dat ze bij beiden elementen geen component van competentie kunnen terugzien.

Klantdoel

Voor het Klantdoel werd aangegeven dat de nieuwe notatie meer intuïtief is. Het resultaat is statistisch niet significant en er kan daarom geen eenduidige conclusie uit deze resultaten worden getrokken. Vijf van de vijftien respondenten geeft aan dat zij beiden elementen gelijk intuïtief vinden. Er zijn verder geen opmerkingen die een duidelijke reden uitspreken voor de voorkeur voor de oude of de nieuwe notatie.

Waardestroom/ Belangrijkheid

Voor de Waardestroom gecombineerd met Belangrijkheid werd aangegeven dat de nieuwe notatie meer intuïtief is. Het resultaat is statistisch niet significant en er kan daarom geen eenduidige conclusie uit deze resultaten worden getrokken. Twee van de vijftien respondenten geven aan niet één van beiden beter te vinden. In de opmerkingen is te zien dat een aantal respondenten de voorkeur geeft aan de kleur voor de belangrijkheid omdat deze de urgentie goed weergeeft. Aan de andere kant wordt er voor de nieuwe notatie gekozen omdat hier de Waardestroom beter wordt weergegeven omdat er een richting wordt aangegeven door de pijl.

5.1.2. Doeltreffendheid

De in 5.1.1 gebruikte juiste antwoorden voor het bepalen van de intuïtiviteit worden met de tijdmeting gecombineerd om op deze manier de doeltreffendheid te bepalen. Op basis van de gegeven antwoorden wordt bekeken of er voor de oude of de nieuwe notatie een betere score wordt gerealiseerd. De resultaten van de vraaggroepen notatieassociatie en begripsvragen zijn statistisch niet significant. Daarom kan er geen waarde worden gehecht aan de uitkomsten van deze vraaggroepen. Voor het onderdeel notatieassociatie (2) kan worden gezegd dat deze statistisch significant is. Op basis hiervan kan worden gezegd dat de oude notatie beter scoort dan de nieuwe notatie in de verhouding juiste antwoorden afgezet tegen de gebruikte tijd.

5.1.3. Reflectie

Naar aanleiding van het uitgevoerde onderzoek kunnen sterke en zwakke punten worden benoemd ten aanzien van de uitvoer van het onderzoek. Als sterke punt kan de ontwikkelde methode om het experiment uit te voeren worden genoemd. Deze methode kan in de toekomst worden gebruikt voor verder onderzoek. Tijdens het experiment zijn de resultaten met betrekking tot de tijd en elementen gedetailleerd verzameld waardoor er voor zowel de oude ten opzichte van de nieuwe notatie een conclusie kan worden getrokken, maar ook meer specifiek op de elementen individueel. Als zwak punt kan het aantal respondenten worden genoemd welke vermoedelijk door de huidige omstandigheden en de specifieke geformuleerde doelgroep lager bleef dan verwacht. Meer specifiek had er daarnaast bij de perceptievragen beter gekozen kunnen worden om de elementen Waardestroom en Belangrijkheid apart te bevragen.

De constructvaliditeit heeft betrekking op het daadwerkelijk meten van wat er beoogd wordt te meten (Saunders, Lewis, & Thornhill, 2016); de intuïtiviteit, en dit kan worden geborgd door diverse soorten vragen te stellen. Dit is in dit onderzoek gebeurd door notatieassociatie, notatieassociatie (2), de begripsvragen en de perceptievragen in het experiment met elkaar te combineren. De interne validiteit heeft betrekking op de mate waarin de bevinding kan worden toegewezen aan de interventie en niet op toeval berust (Saunders, Lewis, & Thornhill, 2016). Dit wordt in dit onderzoek geborgd doordat er respondenten zijn betrokken met een bepaald profiel/ werkzaam in bepaalde sector. Daarnaast hebben zij gedurende het experiment dezelfde informatie ontvangen en is in het experiment nagegaan of zij ervaring hebben met modelleren. De externe validiteit heeft betrekking op de mate waarop het onderzoek generaliseerbaar is. Dit wordt geborgd doordat de respondenten allemaal werkzaam zijn in eenzelfde sector. Het experiment is daarmee te generaliseren voor de IT-medewerkers.

5.2. Conclusies

De intuïtiviteit van een modelleertaal kan worden gemeten door te bepalen welke (variant van een) taal het meest natuurlijk kan worden gebruikt door potentiële gebruikers. Daarbij is van belang dat de gebruiker de modelleertaal snel, juist en zonder veel voorkennis van de taal deze kan gebruiken. De twee PGA-notaties in relatie tot elkaar worden beoordeeld op intuïtiviteit door ze in eenzelfde experiment met elkaar te vergelijken. In dit experiment staat centraal staet welk element (in welke variant) het beste resultaat geeft in juistheid van gebruik en snelheid van gebruik. Door de respondenten diverse typen vragen te laten beantwoorden met betrekking tot de gewijzigde elementen kan worden geconcludeerd of één van de beide notaties inderdaad meer intuïtief is dan andere notatie.

De probleemstelling kan getoetst worden aan de verzamelde onderzoeksresultaten. De probleemstelling die is geformuleerd betreft de volgende: *‘In welke mate en op welke manier leiden de aanpassingen in de PGA-notatie en daarmee de nieuwe notatie tot een significant betere intuïtiviteit dan de oorspronkelijke notatie?’*. Uit de analyse blijkt dat bij toepassing van de nieuwe notatie er sprake is van een statistisch significante hogere intuïtiviteit. Deze resultaten ten gunste van de nieuwe notatie zien we terug in de intuïtiviteit in Notatieassociatie ($p = .002$) en Notatieassociatie (2) ($t(14) = 3.833$, $p = .001$). Uit de analyse blijkt ook dat bij toepassing van de oude notatie er sprake is van een statistisch significante hogere doeltreffendheid. Dit resultaat ten gunste van de oude notatie zien we terug in de doeltreffendheid in Notatieassociatie (2) ($t(14) = -3.344$, $p < .002$).

Ook kan er geconcludeerd worden op basis van de individuele resultaten van de elementen uit Notatieassociatie en Notatieassociatie (2) dat bepaalde elementen van de nieuwe notatie en bepaalde elementen van de oude notatie meer intuïtief zijn. Uit Notatieassociatie blijkt dat vooral Belangrijkheid in de nieuwe notatie meer intuïtief is. Uit Notatieassociatie (2) blijkt dat Competentie, Intern doel, Belangrijkheid, Waardestroom meer intuïtief zijn in de nieuwe notatie. Voor Waardepropositie geldt dat het oude element een betere intuïtiviteit heeft. Het element Klantdoel is hier beide even intuïtief bevonden.

Bij de start van het onderzoek zijn de volgende hypothesen geformuleerd:

Hypothese H_{0_I} : *Er is geen verschil in intuïtiviteit tussen PGA-notatie I en II*

Hypothese H_{a_I} : *De intuïtiviteit van PGA-notatie II is hoger dan die van PGA-notatie I*

Hypothese H_{0_D} : *Er is geen verschil in doeltreffendheid tussen PGA-notatie I en II*

Hypothese H_{a_D} : *De doeltreffendheid van PGA-notatie II is hoger dan die van PGA-notatie I*

Op basis van de resultaten welke in tabel 3 schematisch zijn weergegeven kan worden geconcludeerd dat voor de intuïtiviteit enkel kan worden uitgegaan van de resultaten uit Notatieassociatie en Notatieassociatie (2). De overige resultaten van de Begripsvragen en de Perceptievragen zijn aanwezig, maar zijn statistisch niet significant en de resultaten mogen om deze reden niet betrokken worden. Voor de doeltreffendheid mag enkel worden uitgegaan van Notatieassociatie (2) omdat de Notatieassociatie en de Begripsvragen statistisch niet significant zijn. Op basis hiervan kan worden geconcludeerd:

Hypothese ‘ H_{a_I} ’: *De intuïtiviteit van PGA-notatie II is hoger dan die van PGA-notatie I* is aanvaard. Daarmee kan worden bevestigd dat de nieuwe notatie meer intuïtief is dan de oude notatie.

Hypothese ‘ H_{a_D} ’: *De doeltreffendheid van PGA-notatie II is hoger dan die van PGA-notatie I* is verworpen. Daarmee kan worden ontkracht dat de nieuwe notatie een hogere doeltreffendheid heeft dan de oude notatie.

5.3. Aanbevelingen voor de praktijk

De praktische aanbevelingen die zich vanuit het onderzoek aandoen, hebben in het verlengde van de geformuleerde hypothesen, betrekking op de keuze voor de oude of nieuwe notatie. Wanneer de PGA-notatie wordt ingezet bij besluitvorming waarbij intuïtiviteit een belangrijk aspect is, is het aan te bevelen de nieuwe notatie toe te passen. Besluitvorming op basis van intuïtie, en in het verlengde daarvan de intuïtiviteit van de informatievoorziening voor die besluitvorming, wordt steeds vaker gezien als een goede methode (Burke & Miller, 1999). Het toepassen van de nieuwe notatie kan bij deze besluitvorming daarom dan ook een goede bijdragen leveren aan de besluitvorming. Daarnaast kan de intuïtiviteit zwaarder wegen wanneer modellen worden gepresenteerd aan doelgroepen met geen tot weinig ervaring met de PGA-notatie of modelleren in het algemeen. Wanneer een model wordt gepresenteerd aan doelgroepen met meer ervaring met modelleren, is de intuïtiviteit van ondergeschikt. Zij zijn immers bekend met deze werkwijze. Wanneer een model binnen deze doelgroep wordt ingezet, kan beter worden ingezet op de doeltreffendheid van de notatie. In dat geval is het aan te bevelen om, vanwege de verhoogde doeltreffendheid, de oude notatie te gebruiken.

5.4. Aanbeveling voor verder onderzoek

Dit onderzoek heeft plaatsgevonden met een beperkt aantal respondenten in een specifieke branche en functie. In vervolgonderzoek in een praktische bedrijfscontext zou ervoor gekozen kunnen worden om meer respondenten met een grotere variëteit aan functies en werkzaamheden te laten deelnemen. Hierdoor kan meer inzicht worden verkregen door verschillen tussen bijvoorbeeld opleidingsniveau, aard van de functie, en werkveld te onderzoeken. Een andere mogelijke aanbeveling voor verder onderzoek kan worden gevonden in de complexiteit van de gemodelleerde architectuur. Er kan dan gehypothiseerd worden dat de mate van gepercipieerde intuïtiviteit afhankelijk is van de mate van complexiteit van het model.

Bronnen

- Blazhenkova, O., & Kozhevnikov, M. (2009). The New Object-Spatial-Verbal Cognitive Style Model: Theory and Measurement. *Applied Cognitive Psychology, 23*, 638-663.
- Bork, D., Schrüffer, C., & Karagiannis, D. (2019). Intuitive Understanding of Domain-Specific Modeling Languages: Proposition and Application of an Evaluation Technique. *Conceptual Modeling, 311-319*.
- Burke, L. A., & Miller, M. K. (1999). Taking the mystery out of intuitive decision making. *Academy of Management Perspectives, 13*(4), 91-99.
- Burton-Jones, A., & Meso, P. N. (2006). Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis. *Information Systems Research, 17*(1), 38-60.
- Burton-Jones, A., Wand, Y., & Weber, R. (2009). Guidelines for Empirical Evaluations of Conceptual Modeling Grammars. *J AIS, 10*.
- Dalgarno, M., & Fowler, M. (2008). UML vs. Domain-Specific Languages. *Methods & Tools, 2-9*.
- Figl, K. (2017). Comprehension of Procedural Visual Business Process Models. *Business & Information Systems Engineering, 59*(1), 41-67.
- Figl, K., & Recker, J. (2016). Exploring cognitive style and task-specific preferences for process representations. *Requirements Engineering, 21*(1), 63-85.
- Gemino, A., & Wand, Y. (2004). A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering, 9*(4), 248-260.
- Gemino, A., & Wand, Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering, 55*(3), 301-326.
- Grandhi, S., Joue, G., & Mittelberg, I. (2011). *Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces*.
- Lankhorst, M. e. a. (2009). *Enterprise Architecture at Work*. Berlin Heidelberg: Springer Verlag.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics, 2*(1), 21-33.
- Roelens, B., & Bork, D. (2020). *An Evaluation of the Intuitiveness of the PGA Modeling Language Notation*. Cham: Springer.
- Saunders, M. N. K., Lewis, P., & Thornhill, A. (2016). *Research methods for business students*.
- Shanks, G., Gloet, M., Asadi Someh, I., Frampton, K., & Tamm, T. (2018). Achieving benefits with enterprise architecture. *The Journal of Strategic Information Systems, 27*.
- Tamm, T., Seddon, P. B., Shanks, G., & Reynolds, P. (2011). How does enterprise architecture add value to. *Communications of the Association for Information Systems, 28*(1), 141-168.
- Thurmond, V. A. (2001). The Point of Triangulation. *Journal of nursing scholarship, 33*(3), 253-258.
- Turner, P. (2008). Towards an account of intuitiveness. *Behaviour & Information Technology, 27*(6), 475-482.
- Wand, Y., & Weber, R. (2002). Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda. *Information Systems Research, 13*(4), 363-376.

Bijlage 1: Literatuuranalyse

In het artikel van Gemino en Wand (2005) zijn twee versies van het *entity-relationship model* (ERM) vergeleken (*Intra-Grammar Comparison*). Zij schrijven dat het model met de optionele onderdelen in eerste instantie minder complex lijkt, terwijl het model met de verplichte onderdelen meer complex oogt. Naar aanleiding van het onderzoek blijkt dat het model met de verplichte onderdelen een meer heldere voorstelling geeft. Deze oogt meer complex, maar is minder dubbelzinnig en dus beter te begrijpen door de gebruiker. Dit heeft ermee te maken dat gebruikers het model in 'stukjes' analyseren en opnemen en niet als één groot geheel. Hierdoor kan er meer gedetailleerde informatie worden opgenomen, waardoor in het hoofd van de gebruiker het model beetje bij beetje wordt opgebouwd en er een gedetailleerd geheel ontstaat (Gemino & Wand, 2005).

De paper van Burton-Jones en Meso (2006) evalueert het *good decomposition model* (GDM) om na te gaan in welke mate conceptuele modellen de betekenis over een domein communiceren aan analisten. 'Het *decomposition-process* zorgt ervoor dat een warrig domein uit de normale wereld (volgens de opvattingen van de gebruikers van dat domein) in kleinere onderdelen wordt onderverdeeld om zo structuur en gedrag te achterhalen.' (Burton-Jones & Meso, 2006). GDM geeft vijf voorwaarden (i.e. *minimality, determinism, losslessness, weak coupling, en strong cohesion*) die nodig zijn om een domein te ontleden waardoor er een goede communicatie kan worden bewerkstelligd. Conceptuele modellen hebben volgens Burton-Jones en Meso (2006) ten minste vier doelen: 1. het begrip van de analist ten aanzien van het domein verbeteren 2. ontwikkelaars en gebruikers te helpen communiceren 3. als basis van het ontwerp van hetgeen ontworpen is dienen 4. als documentatie van de oorspronkelijke vereisten van het systeem fungeren. In deze paper wordt gezocht naar de elementen die van invloed zijn op de mate waarin conceptuele modellen met een bepaalde notatie betekenis overdragen aan analisten.

Figl (2017) heeft in het onderzoek als doel om te verhelderen welke bronnen van cognitieve inspanning nodig zijn om procesmodellen te begrijpen en om inzicht te krijgen in welke factoren de intuïtiviteit en de begrijpbaarheid beïnvloeden. Zij gaat daarmee op zoek naar factoren die de visuele procesmodellen beïnvloeden. Volgens haar heeft een verbeterde begrijpelijkheid een directe invloed op de ontwikkeling, efficiëntie en kosten van informatiesystemen en naast een betere bruikbaarheid voor de gebruikers en de systeembeheerders zorgt het ook voor een betere kwaliteit in het model. Cognitieve capaciteit kan een probleem vormen in de begrijpbaarheid van een model (Figl, 2017).

'Procesmodellen omschrijven iemands begrip van processen' schrijven Figl en Recker (2016). Processen kunnen op verschillende wijzen worden omschreven. Afhankelijk van het doel waarvoor ze gebruikt gaan worden, wordt een aanname gemaakt welke wijze de beste is. In de paper van Figl en Recker (2016) wordt onderzocht of deze aanname juist is. Er wordt empirisch onderzocht wat de voorkeur is van een gebruiker voor verschillende procesrepresentatievormen afhankelijk van het doel en zijn/haar cognitieve stijl. Verschillende gebruikers hebben over het algemeen een verschillende voorkeur qua procesweergave. Om deze reden onderzoeken zij de individuele cognitieve stijl in relatie tot voorkeuren voor procesweergave. Daarnaast onderzoeken zij of het de voorkeur geniet om op basis van het doel een bepaalde weergave te kiezen, dus voor een specifieke taak een specifieke weergave.

Volgens Gemino en Wand (2004) hebben conceptuele modellen een viertal rollen in het ontwikkelen van domeinbegrip: 1. iemands redeneervermogen over een domein helpen verhogen, 2. het communiceren van domeingegevens tussen belanghebbenden, 3. het communiceren van domeingegevens tussen systeemontwerpers, 4. het documenteren van het domein voor toekomstig gebruik. Samenvattend kan daarmee worden gezegd dat conceptueel modelleren kan worden gezien als een proces waarbij individuen redeneren en communiceren over een domein om zo het begrip van dit domein te verbeteren (Gemino & Wand, 2004). Om conceptuele modellen te kunnen vergelijken is het belangrijk te weten wat er wordt vergeleken. Als we *scripts* vergelijken, bieden we vergelijkende informatie over een enkel

gemodelleerd domein. Als we grammatica vergelijken dan zoeken we vergelijkende informatie over het vermogen van de grammatica om elk domein te modelleren. *Conceptual modeling techniques* (CMT) focussen op het modelleren van de domeinen in de 'echte' wereld. Ontologie is een verzameling concepten en premissen over wat er kan bestaan en gebeuren in een bepaald domein of in de wereld. Ontologie houdt zich bezig met de manier waarop mensen de wereld om hen heen beschrijven. Gemino en Wand (2004) omschrijven dat in eerdere studies CMT's zijn vergeleken. De eerste twee onderzoeken die zijn omschreven zijn gericht op het diagram zelf en op de informatie die hierdoor wordt overgebracht. In later onderzoek wordt de focus verlegd van het bestuderen van het product naar het proces van het begrijpen van het model waarbij cognitieve activiteit het onderwerp is. Meer recent is er door onderzoekers onderzocht op basis van probleemoplossingstaken, wat ervoor zorgt dat de respondenten gaan redeneren over het domein dat wordt gepresenteerd. Het verschil tussen de eerste onderzoeken en het laatste onderzoek zit in de verschuiving van het begrijpen van het diagram zelf naar het mentale model van het domein dat wordt ontwikkeld door het diagram te gebruiken. De meting van het begrip wordt gedaan door de deelnemers te laten redeneren over het domein dat wordt beschreven, in plaats van de elementen in een model te herkennen. Conclusie aan de hand van het voorgaande is: 1. sluitende empirische vergelijkingen van CMT's zijn moeilijk te maken, 2. de diversiteit aan constructen en procedures suggereert dat er geen algemeen aanvaarde instrumenten, procedures of constructen voor vergelijkende prestaties bestaan. De focus van de observatie geeft aan of de meting is gericht op het gebruiksproces van de CMT of het resultaat van het gebruik van de CMT. Gemino en Wand (2004) hebben aan de hand van bovenstaande conclusies een *framework* ontwikkeld om conceptuele modellen te vergelijken. Dit *framework* geeft een aantal dimensies om bestaand werk te vergelijken, te beoordelen waar *gaps* zitten en om de bepalen waar het model verder ontwikkeld kan worden.

Burton-Jones, Wand en Weber (2009) omschrijven dat er modellen bestaan om conceptuele modellen te evalueren maar dat er nog onduidelijkheden zijn die opgehelderd dienen te worden: 1. prestatiecriteria die kunnen worden gebruikt om conceptuele modellering grammatica's te evalueren, 2. karakteristieken van de grammatica die deze performance beïnvloeden, 3. factoren die onderzoekers zouden moeten overwegen als zij het effect van grammaticakarakteristieken op prestatie van de grammatica gaan testen. Burton-Jones, Wand en Weber (2009) hebben in hun onderzoek richtlijnen ontwikkeld om de eerdergenoemde drie onduidelijkheden te gaan ophelderen en daarnaast geven zij een aanvulling op eerder gedane studies. De studie van *semantics* focust op de betekenis van woorden en zinnen in een taal. De studie van *pragmatics* focust op hoe talen in de praktijk worden gebruikt. De studie van *pragmatics* kan worden verdeeld in de *denotational* en *connotational* betekenis. De studie van *denotational* focust op de objectieve relatie tussen woorden en zinnen. De *connotational* focust op hoe mensen betekenis geven aan woorden en zinnen die vaak afwijken van de objectieve betekenis.

Wand en Weber (2002) ontwikkelen in hun paper een *framework* om de volgende vraag te kunnen beantwoorden: 'hoe kunnen we de wereld modelleren om onze ontwikkeling, implementatie, gebruik en onderhoud van waardevollere informatiesystemen te vergemakkelijken'. Hun doel is onderzoeksonderwerpen te vinden die dit probleem erkennen en zo een verder uiteenzetting te kunnen maken. Zij geven aan dat het vooral interessant is om de effectiviteit en de efficiency te onderzoeken in de toekomst.

Naast het modelleren van de talen heeft dit onderzoek ook betrekking op de intuïtiviteit van de modelleertaal. Turner (2008) omschrijft intuïtiviteit in twee verschillende betekenissen. Intuïtiviteit gebaseerd op vertrouwdheid (vergeleken met *know-how*) en intuïtiviteit die onze uitdrukking weerspiegelt. In zijn paper onderzoekt hij de meest wenselijke eigenschap van intuïtieve *human-computer interactie*. Intuïtiviteit werd jarenlang geïnterpreteerd als gebruiksgemak, of leergemak, consistentie of het vertrouwd zijn. Er is echter bij ontwerp geen eenduidige definitie gegeven waar de systemen of apparaten dan aan moeten voldoen om intuïtief te zijn. Er worden dan ook verschillende eigenschappen voor intuïtiviteit gehanteerd. Doel van de paper is om intuïtiviteit te deconstrueren. Intuïtiviteit zou

wellicht beter werken omdat het uitgaat van reeds bestaande routines en sociale *know-how*. Dit zorgt voor een lastenverlichting in de hersenen doordat je werkt met zaken die al bekend zijn voor de gebruiker.

Intuïtiviteit is eerder onderzocht door Grandhi, Joue en Mittelberg (2011). Zij onderzoeken in hun paper de intuïtiviteit van bepaalde gebaren. Zij hebben daarbij een experiment opgezet om de intuïtiviteit van gebaren te bepalen. Zij hebben daarbij de deelnemers een aantal acties die uitgebeeld zijn op papier op drie verschillende manieren te laten uitbeelden. Conclusie van hun onderzoek is dat: 1. gebaren die een actie teweeg moeten brengen zouden de beweging moeten laten zien in plaats van een statisch symbool. 2. Voor gebaren die een taak moeten uitbeelden waarbij bepaald gereedschap gebruikt dient te worden zou je geen *body-part-as-tool*-gebaar voor moeten gebruiken omdat dit niet natuurlijk voelt en niet intuïtief is. Bijvoorbeeld als je een object wil snijden kan je beter doen alsof je hand een mes vastheeft dan te doen alsof je hand het mes is en daarmee te snijden. 3. gebaren om het aanpassen van objecten over te brengen, moeten met twee handen worden uitgevoerd. De niet-dominante hand (in het onderzoek bij (bijna) alle deelnemers links) voert daarbij meestal het onderwerp uit en de dominante hand (in het onderzoek bij (bijna) alle deelnemers rechts) voert de actie uit.

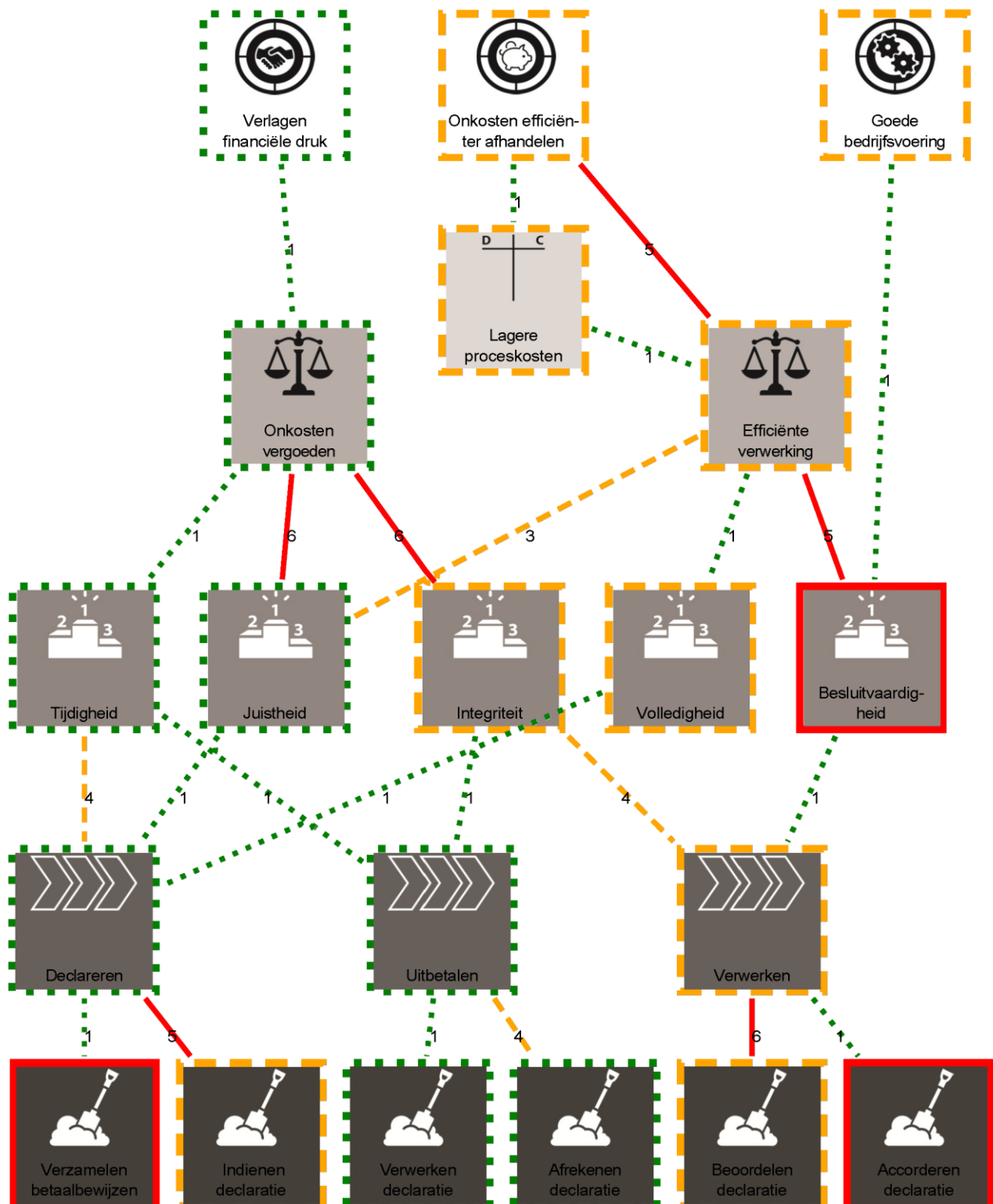
Bijlage 2: Vragenlijst en gebruikte modellen in de enquête.

Deze bijlage is als PDF-bestand ingevoegd.

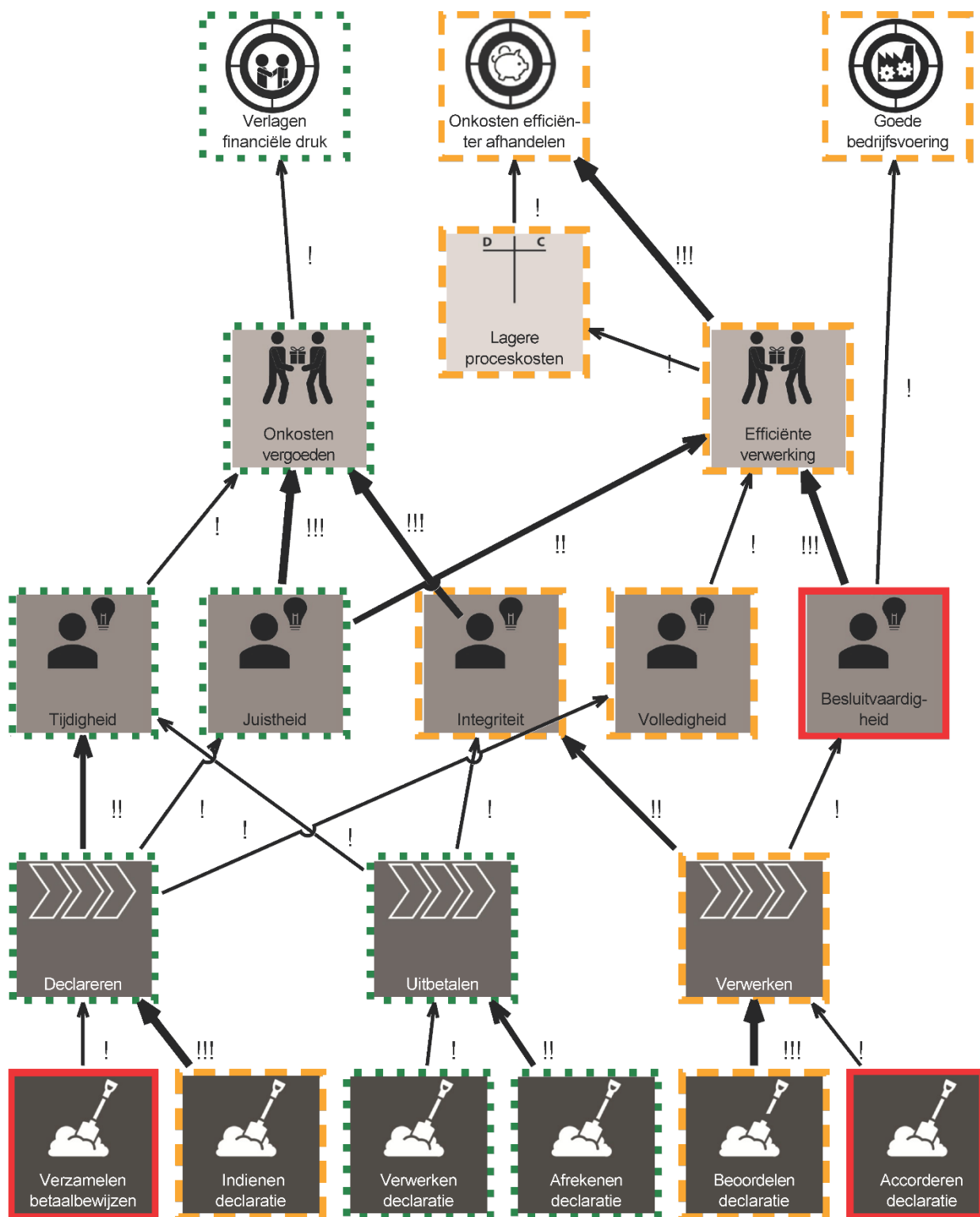


Limesurvey
experiment.pdf

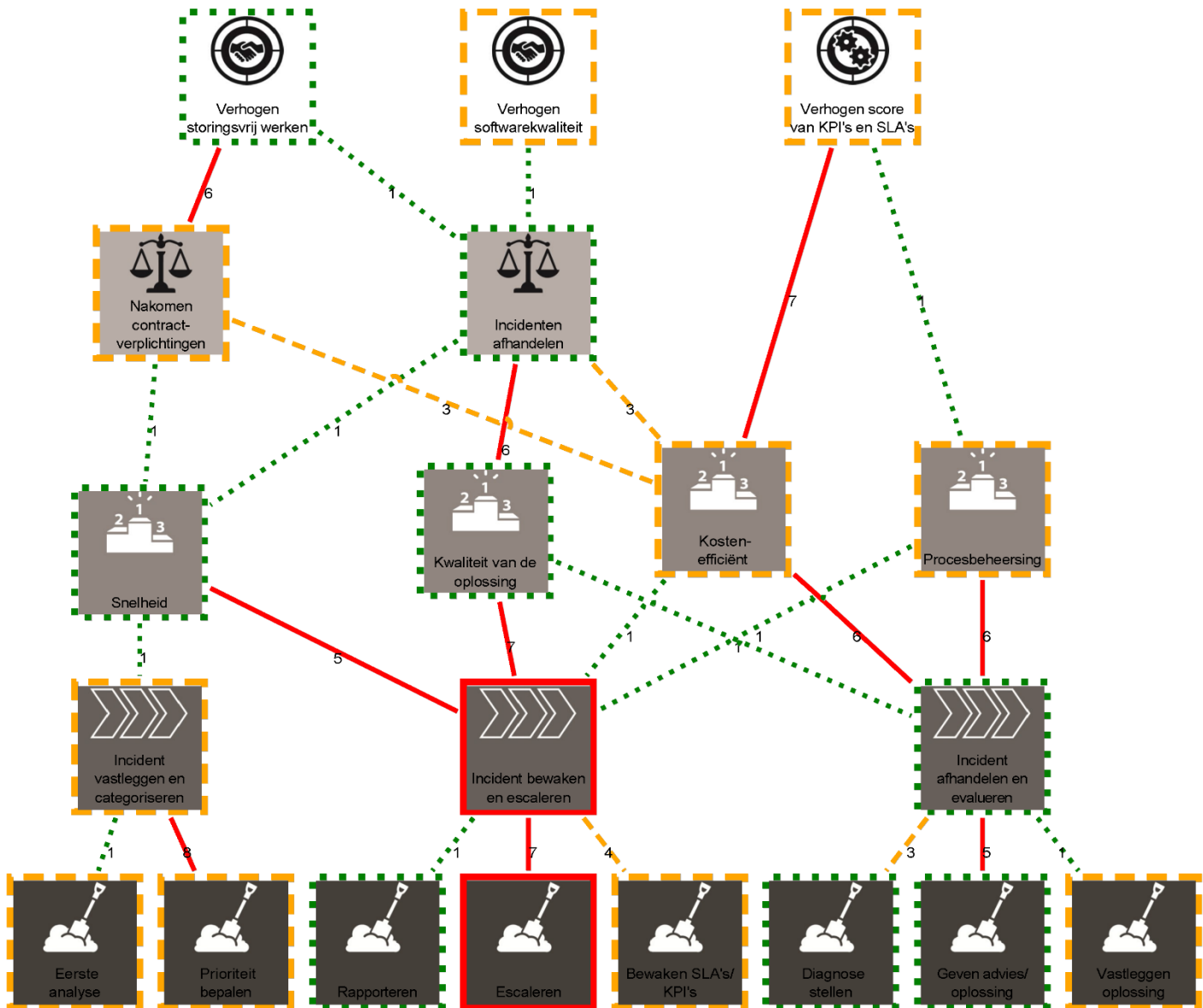
Bijlage 3: Bedrijfsarchitectuur begripsvragen



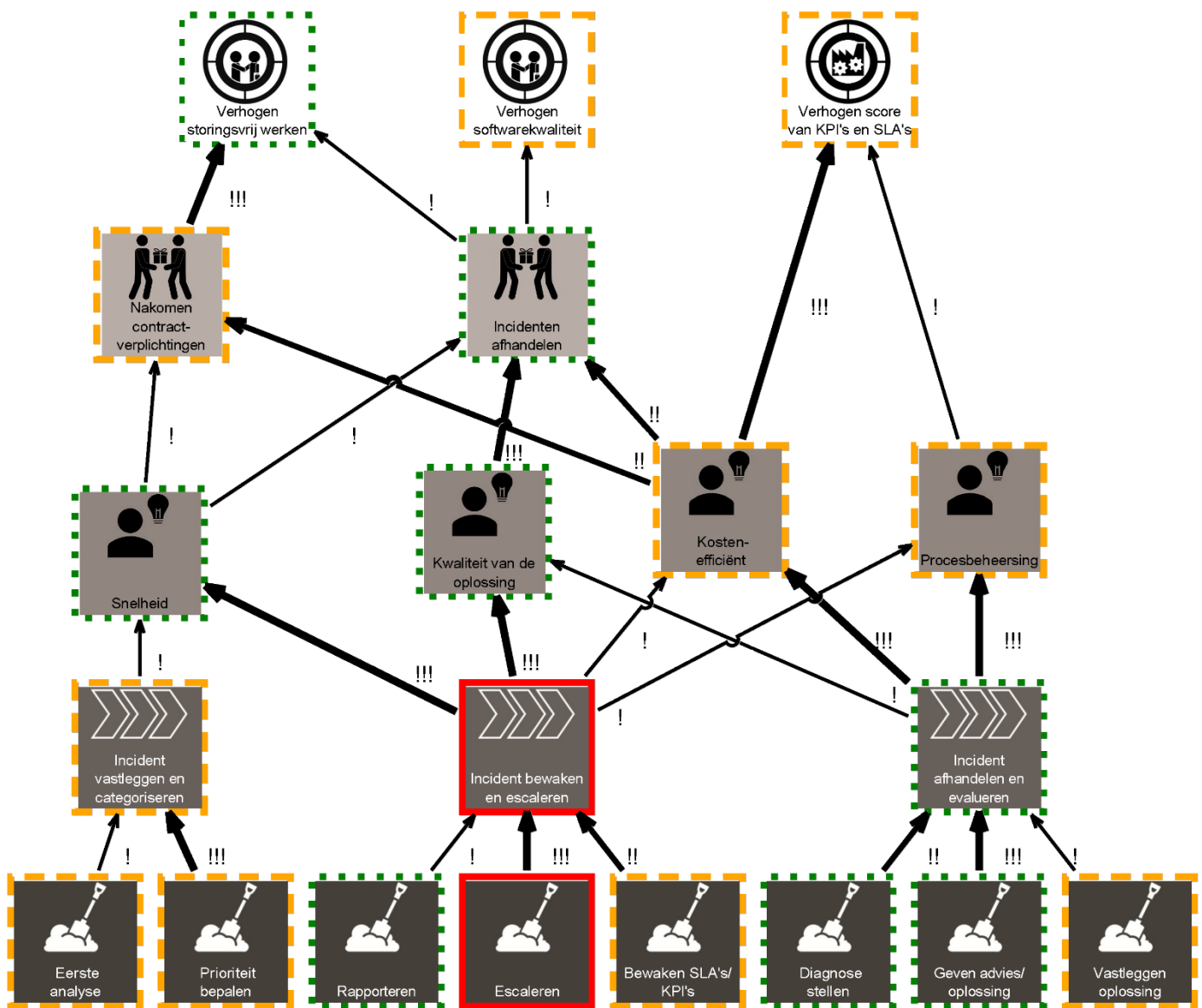
Figuur 1: Model 1 oude notatie



Figuur 2: Model 1 nieuwe notatie



Figuur 3: Model 2 oude notatie



Figuur 4: Model 2 nieuwe notatie

Bijlage 4: Data-bestand

Deze bijlage is als excel-bestand ingevoegd.



YT_dataset.xlsx